DNA-BARCODING OF NON-BITING MIDGE LARVAE (CHIRONOMIDAE) IN HEAVY METAL POLLUTED PONDS

By

Valeria Manfredi

Canterbury Christ Church University

Thesis submitted for the degree of MSc by Research

2019

Table of Contents

List of Figure and Tables

Abstract

Chironomids, or non-biting midges (Chironomidae), belong to the order of true flies (Diptera), and are frequently used as bio-indicators of water quality, where higher diversity levels reflect less polluted habitats. However, the identification of chironomids (as well as other freshwater invertebrates) is done using morphological characters, a process that can be difficult and time-consuming. With the use of the taxonomic identification tool DNA barcoding, it is possible to identify the specimens to genus or species level without the use of morphological identifications which often difficult and time-consuming. In this study, DNA barcoding using the mitochondrial (mt) DNA the cytochrome oxidase 1 gene (CO1) was used to identify chironomid larvae from seven different ponds in Wales in relationship to sediment and water environmental characteristics. Chironomid larvae, sediment and water samples were taken from five ponds in Parys Mountain, an old copper mine in northern Wales, and from two ponds in Cwm-Idwal in Snowdonia Natural Park. Because of the history of the mining site, four ponds from Parys Mountain were suspected to be polluted. Environmental conditions (pH, dissolved oxygen, conductivity, salinity and temperature), magnetic susceptibility (MS) and isothermal remanent magnetisation (IRM), and presence of heavy metals were analysed to characterize the sites and to understand the effect on chironomid larvae diversity and community structure. Chironomid larvae were DNA-barcoded and the genetic and community structure for each pond were calculated and analysed in relationship with environmental conditions, magnetic characteristics and presence of heavy metals. A total of 15 genera of chironomids were found among the seven ponds. The results indicated a significant difference between suspected non-polluted ponds (WAPM1,WACI1,WACI2) and most of the suspected polluted ponds (WAPM2,WAPM3,WAPM4,WAPM5); MS and IRM analyses showed no significant differences among the seven ponds ($p=0.206$ and $p=0.653$, respectively). Results of the analysis of both the sediments and water samples showed a higher concentration of heavy metals in the suspected polluted ponds than in the suspected non-polluted ones. However, a significant

difference was found in the community structure of chironomid larvae among the ponds by using ANOSIM analysis ($p= 0.001$). The suspected polluted ponds had in lower diversity compared to the suspected non-polluted ones. In conclusion, there seems to be a correlation between the chironomid larvae diversity and community structure with the environmental composition of the ponds. This study represents the first DNA barcoding study of chironomids in polluted and non-polluted ponds in Wales and show that this technique is an efficient way for estimating insect diversity and community structure in relation with water quality biomonitoring.

Introduction

WATER QUALITY ASSESMENT

Water is one of the primary natural resources on earth (Almeida et al. 2008). Accepting the importance of water for maintaining life, is a matter of highest environmental concern (Poonam et al. 2013). "Water quality" is a term used to express the suitability of water to sustain various processes. The Conventional methods for evaluating quality of water are based on the comparison of experimentally determined parameter values with the existing guidelines (Debels et al. 2005). Therefore, water quality indices are approaches that minimizes the data volume and simplifies the expression of water quality status.

Water quality is affected by a wide range of both natural (geological; hydrological and climatic) and anthropogenic (municipal and industrial wastewater discharge) inputs (Bartram, 1996). According to Kazi et al. (2009), human activities are a major factor determining the quality of both the surface and groundwater through atmospheric pollution, effluent discharges, agricultural chemicals and land use. Pollution of the aquatic environment, (GESAMP 1988), occurs when humans introduce, either by direct discharge to water or indirectly (for example through atmospheric pollution or water management practices), substances or energy that result in deleterious.

As Bartam (1996) states, the quality of the aquatic environment is a broader issue which can be described in terms of: Water quality; the composition and state of the biological life present in the water body and the nature of the particulate matter present, and the physical description of the water body (hydrology, dimensions, nature of lake bottom or river bed, etc.).

The complete assessment of the quality of the aquatic environment, therefore, requires that water quality, biological life, particulate matter and the physical characteristics of the water body be investigated and evaluated. This can be achieved through: the chemical analyses of water, particulate

matter and aquatic organisms (such as planktonic algae and selected parts of organisms such as fish muscle); biological tests, such as toxicity tests and measurements of enzyme activities; the descriptions of aquatic organisms, including their occurrence, density, biomass, physiology and diversity (from which, for example, a biotic index may be developed, or microbiological characteristics determined) and the physical measurements of water temperature, pH, conductivity, light penetration, particle size of suspended and deposited material and dimensions of the water body.

ENVIRONMENTAL POLLUTION

The natural environment comprehend all living and non-living things occurring in nature. It consists of the flora, fauna and the abiotic, and includes the aquatic, terrestrial and atmospheric habitats (Johnson et al., 1997). In contrast to the natural environment is the artificial environment which comprehend the human transformation of landscapes as urban settings (Johnson et al., 1997). A pollutant is any substance in the environment, present beyond a set or tolerance limit, which causes objectionable effects to the environment, impairing the welfare of the organisms living there, reducing the quality of life and may eventually cause death to the flora and fauna.

A pollutant is typically transformed into end products different than the chemical form in which it was initially emitted (Hill, 2010). When large quantities of rock containing sulphide mineralsare excavated from an open pit or opened up in an underground mine, it reacts with water andoxygen to create sulphuric acid. When the water reaches a certain level of acidity, a naturally occurring type of bacteria, *Thiobacillus ferroxidans*, may undergo population growth, accelerating the oxidationand acidification processes, leaching even more trace metals from the wastes. The acid will leach from the rock as long as its source rock is exposed to air and water and until the sulphides are leached out a process that can last hundreds, even thousands of years. Acid is carried off the minesite by rainwater or surface drainage and deposited into nearby streams, rivers, lakes and groundwater. This acid drainage from mines severely degrades water quality, and

can kill aquatic life and make water virtually unusable (Environmental Mining Council of BC. 2000).

Human activities have introduced pollutants of different kinds into the environment all over the world, for example: air pollutants from industry, exhaust emissions from transportation vehicles, radionuclides from nuclear weapons tests and uranium mill tailings, pesticides from agricultural fields, sewage from domestic wastewater, heavy metals, and other chemicals that enter lakes, rivers, surface water, and groundwater (Gilbert, 1987). Also, some environmental factors such as pH, dissolved oxygen, salinity, temperature and conductivity can influence biodiversity because different species are able to survive and cope at different ranges of each of them (Panatta et al. 2007; Kranzfelder et al. 2017). Particularly important are heavy metal pollutants introduced into the environment by mining processes, which could have an impact on biodiversity. Chemical contamination of the environment, including the effects of long- term, low-level chronic exposure of populations, as well as short-term, acute exposures like oil spills, has been classified as one of the major causes of biodiversity loss and it has been implicated in the decline or disappearance of many animal species (Pimm et al. 1995).

Biodiversity may be defined as the number, variety, and variability of living organisms within a temporal and spatial scale. It exists as a continuum, including higher taxa genera, families, etc., species, populations, subpopulations, individuals, and genes. Disruption of genetic equilibria at any of these levels has a direct bearing on the decline of diversity, subsequent enhancement of vulnerability to environmental stress, and extinction of species. Both extinction and the evolution of new species have been an integral part of the history of life on earth. However, due to the accelerated destruction of natural habitats, intensification of agriculture, and chemical pollution, the present extinction rate is estimated to be 10–100 times the historical background (Bickham et al. 2000).

Pollutants stay at the point of release. They move and are transported, among air, water, soil, and sediments, and often food as well (Hill, 2010). They often move transboundary: across state and national boundaries traveling with air or water currents (Hill, 2010). Moreover, biotransport occurs when pollutants are carried in body tissues of migrating animals such as salmon, whales, or birds, or

are found in the droppings of migratory birds. Many monitoring and research studies are currently being conducted to quantify the amount of pollutants entering the environment and to monitor ambient levels for trends. Other studies seek to determine how pollutants disperse and persist in air, water, soil and biota, and to determine the effects of pollutants on humans and the environment (Gilbert, 1987).

PARYS MOUNTAIN AND POLLUTION

The United Kingdom has a long history of mining for metals, dating back at least 4000 years, which has produced a vast number of mines, with over 3700 sites in Wales, the South West of England and Northumbria alone (Marsay, 2018). Wales, is particularly rich in mineral resources, the type and distribution of which are related to the complex geologic and tectonic history (Marsay, 2018). Therefore, Parys Mountain (Wales) site has been chosen for the sampling analysis in this research.

The Parys Mountain copper–lead–zinc deposit of Anglesey (North Wales) represents a volcanogenic massive sulphide district of major metallogenic importance, which is characterized by the occurrence of concordant massive to banded sulphide lens formed by volcanic processes normally on the sea floor (Rios et al. 2008). The mines of Pays Mountain constitute an industrial archeological monument of international importance (Wilson et al. 2007). However, beside the historical and archaeological importance, the mines pose several risks. These are primarily related to the water environment as a consequence of sulphide oxidation and acidic drainage (Rees, 2005). Parys Mountain copper mines on Anglesey are the single largest contributor of copper and zinc to the Irish Sea, discharging 24 tons of zinc and 10 tons of copper every year (Environment Agency, 2003).

Parys Mountain precipitation ponds were used to extract copper from water. It was discovered that purer metal could be obtained very efficiently by precipitation from solution. Water was pumped to the top of the mountain and allowed to drain down through the spoil and the underground workings, dissolving the copper due to its very acid nature. Scrap iron was then added to the copper-rich water to give metallic copper in a sequence of purpose-built, brick-lined "precipitation ponds" of which

there are several examples on the mountain, the best preserved being those in this central valley. The dissolved iron was itself then oxidised and precipitated as ochre, a valuable by-product that was marketed as a pigment. (Anglesey Mining plc.). Moreover, copper, zinc and lead were suspected to be present in the site (Karataglis, 1982; Wilson et al. 2006). Rocks from Parys Mountain originated as muds about 440 million years ago by erupting lavas, ashes and fumes from submarines volcanoes that by depositing on the sea floor produced rich deposit of metals which is unique in Britain (Anglesey Mining plc.). These metals occur as pyrite (iron), sulphide minerals (copper and iron), galena (lead) and sphalerite (zinc) (Anglesey Mining plc.; Karataglis, 1982; Wilson et al. 2006).

Previous monitoring has shown that some abandoned metal mines are significant contributors to heavy metal pollution in rivers and seas through acid mine drainage (AMD) which can have many effects on the environment. AMD is the product of water coming into contact with sulphide minerals and being exposed to the atmosphere, and it occurs at almost all mines that have sulphide deposits (Marsay, 2018). AMD, lowers the pH of the water makings the environment uninhabitable for many flora and fauna, and those that can survive in the lower pH are threatened by toxic metals, which are more bio-available in low pH environments (Marsay, 2018), AMD is present in Parys Mountain.

Environmental pollution by heavy metals is very prominent in the mining areas (Peplow, 1999). These metals are leached out and in sloppy areas, are carried by acid water downstream or run-off to the sea (Duruibe et al. 2007). Through mining activities, water bodies are most emphatically polluted (Garbarino et al. 1995). The potential for contamination is increased when mining exposes metal-bearing ores rather than natural exposure of ore bodies through erosion (Garbarino et al. 1995), and when mined ores are dumped on the earth surfaces in manual dressing processes. Through rivers and streams, the metals are trans- ported as either dissolved species in water or as an integral part of suspended sediments, (dissolved species in water have the greatest potential of causing the most deleterious effects). They may then be stored in river bed sediments or seep into the underground water thereby contaminating water from underground sources, particularly wells; and the extent of contamination will depend on the nearness of the well to the mining site (Duruibe et al. 2007).

HEAVY METALS IN THE SEDIMENTS

The term “heavy metals” refers to any metallic element that has a relatively high density and is toxic or poisonous even at low concentration (Lenntech, 2004). “Heavy metals” is a general collective term, which applies to the group of metals and metalloids with atomic density greater than 4 g/cm^3 or five times or more, greater than water (Huton and Symon, 1986; Battarbee et al. 1988). However, the definition of heavy metal has little to do with density but concerns the chemical properties of the elements. Heavy metals include lead (Pb), cadmium (Cd), zinc (Zn), mercury (Hg), arsenic (As), silver (Ag) chromium (Cr), copper (Cu) iron (Fe), and the platinum group elements.

Heavy metals entering the aquatic environment partly become associated with fine-grained particulates and, as a result of settling accumulate in sediments (Farkas et al. 2007). Inorganic and organic nutrients are, in varying amounts and forms, almost continuously transported to lake bottoms by sedimentation. Due to various biological, physical, chemical and mechanical processes nutrients can be returned to the free water from the sediments. This cycling between lake sediments and water may occur according to various schemes dependent on lake type and bottom conditions. The rate of nutrient turnover will be strongly influenced by hydrological conditions, lake morphology, water residence time, temperature regimes, and size and density of particles.

Magnetic susceptibility (MS) was proven to help in probing the mineralogy concentration and grain-size distribution of several mineral in sediments (Yang et al., 2016). Sediment magnetic susceptibility has proved to be an indicator of soil properties regarding various environmental conditions (Lourenco et al., 2014) as it is influenced by climate change and anthropogenic activities such as industrial pollution, burning of fossil fuels and agricultural activities (Lu et al., 2012). In the sediments, metal ions have been found to partition into different chemical forms associated with a

variety of organic and inorganic phases, depending on chemical and geological conditions (Chao, 1984). Many studies have demonstrated the remobilization (release) of contaminants from bed sediments, which can thus elicit acute toxicity but, more frequently, chronic and sublethal responses in aquatic organisms. Moreover, as the direct uptake from interstitial water and ingested sediment threatens infaunal and demersal organisms, the trophic transfer of persistent and accumulating pollutants can, in turn, extend the hazard from the sediment compartment to the entire community of the aquatic ecosystem. Such sediment potential for being a sink as well as a source of contaminants can make sediment chemistry and toxicity key components of aquatic system quality. This could be particularly useful to investigate large riverine environments where the aqueous concentrations of pollutants are frequently close to the detection limits of chemical and toxicological methods. Moreover, such concentrations are often highly variable because of several interacting factors such as source characteristics, flow regimes of tributaries and receiving waters, as well as their mixing dynamics (Arillo et al. 2003).

BENTHIC INVERTEBRATES

Benthic macroinvertebrates [the term “benthic” means “bottom-living”, referring to organisms that usually inhabit bottom substrates for at least part of their life cycle (Rosenberg and Resh, 1993)] are commonly used in water quality assessments (Kranzfelder et a*l*. 2017). Benthic invertebrates are common inhabitants of lakes and streams where they are important in moving energy through food webs. The term "benthic" means "bottom-living", so these organisms usually inhabit bottom substrates for at least part of their life cycle (Rosenberg and Resh, 1993).

The most diverse group of freshwater benthic invertebrates is the aquatic insects. Thus, as a highly diverse group, benthic macroinvertebrates are excellent candidates for studying of changes in biodiversity driven by both abiotic and biotic factors (Raunio et al. 2011). However, benthic macroinvertebrates can be difficult to work with unless the proper study design is used (Rosenberg

and Resh, 1993). For example the quantitative sampling is difficult because of the distribution of benthic macroinvertebrates requires large numbers of samples to achieve reasonable precision in estimating population abundance. Also, the resulting processing and morphological identification requirements for samples can be costly and time consuming as some groups of benthic macroinvertebrates are taxonomically difficult and the distribution and abundance of benthic macroinvertebrates are affected by a large number of natural factors, which have to be accounted for to determine changes in biodiversity (Rosenberg and Resh, 1993).

The collection of benthic macroinvertebrates from lakes and streams is usually a straightforward procedure using standard equipment. However, the removal of organisms from background material can be tedious and time-consuming unless available labor-saving strategies are used and the identification of organisms to the species level, when possible, requires substantial training and skill. The processing of samples can be successfully accomplished by non-specialists, but the involvement of systematists is recommended for species-level identifications (Mandaville, 1997).

CHIRONOMIDAE

Among benthic macroinvertebrates, the family Chironomidae (Diptera), is mainly used as bio-indicator of water quality (Carew et al. 2007; Punti et al. 2009). The family Chironomidae, commonly known as chironomids or non-biting midges, is a large and diverse family of true flies, and are globally distributed among most types of freshwater ecosystems including rivers, streams, lake, ponds and wetlands (Lee et al. 2006; Punti et al. 2009; Panatta et al. 2007). There are over 10,000 species known worldwide (Armitage, cited in Ekrem et al. 2010). Non-biting midges are important in the aquatic food chain because they are one of the major food sources for other invertebrates and vertebrates (Cranston, cited in Lee et al. 2006). Chironomid larvae are elongated and cylindrical and their size usually vary from 5 to 10 mm according to larval stage and environmental factors (Dillon, 1985). They have a characteristically sclerotized head capsule that is used for morphological

identification, together with the last part of the segmented body ("tail") and their pair of unjointed "prolegs" in the first segment of the thorax (Fig. 1).

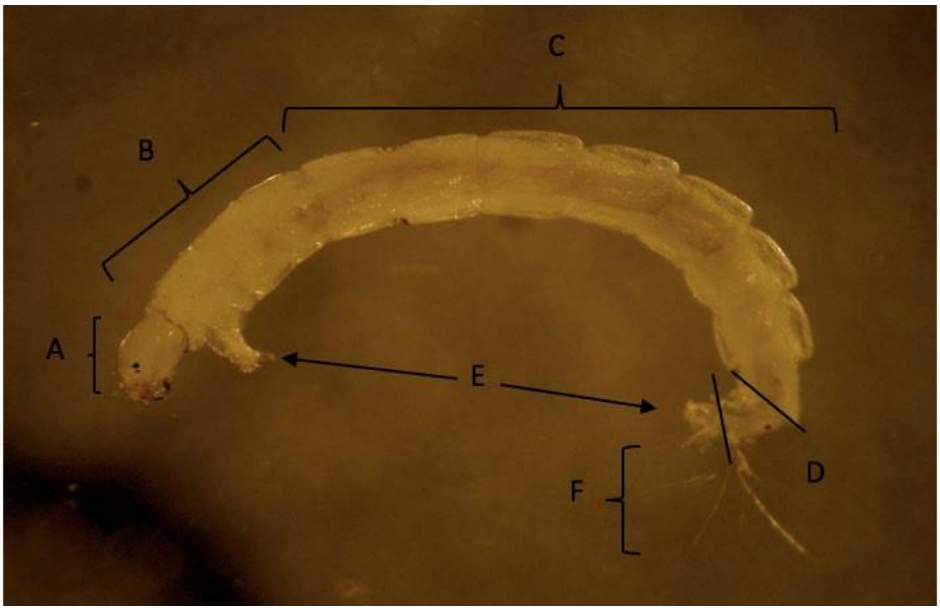


Figure 1. (30X magnification) Unidentified chironomid larva (Chironomidae) showing segmentation typical of arthropods. A: head capsule; B: thorax: C: abdomen (10 segments, including the two posterior segments); D: two posterior segments; E: prolegs or parapods; F: procercus. The 13 segments of thorax plus abdomen are loosely refered as abdomen (Cranston, 1982).

Chironomid larvae vary in colour. Most of them are tan or brown, some of them are white or green, but just larvae from the subfamily Chironominae are pink or red because they have haemoglobin in their fluids that help them cope with low oxygen level, thus they are commonly known as "blood-midges" (Foote, 1987).

Many factors, such as organic content of sediments, pH, dissolved oxygen, salinity and aquatic vegetation, can affect composition and distribution of Chironomid larvae (Panatta et al. 2007; Kranzfelder et al. 2017). Chirnomid larvae can be used as bio-indicators of water quality because they are commonly found in most aquatic systems (Kranzfelder et al. 2017; Martin et al. 2008; Panatta et al. 2007), they are usually sedentary and represent local conditions, they are benthic and thus closely related to sediments (Carew et al. 2007), and they accumulate heavy metals in a way directly proportional to environmental conditions (Beaty et al. 1998). Sediments have an important role in Chironomid larvae life (Pinder, 1986; Armitage et al. 1995). Sediments are mainly a food sources, a

cover from predators and a source of potential toxic compounds (De Haas et al. 2006). Animals used in water quality assessment studies, should be identified to the lowest possible taxonomical level to avoid confusing in contaminant concentrations among sites and over time because of behavioural and physiological differences among taxa (Rainbow, 2002; Skubala et al. 2004). However, identifying freshwater invertebrates to the species level is often difficult.

Chirnomid larvae can be identified morphologically or through the use of DNA barcoding. Morphological identification is time-consuming, and it requires taxonomic expertise (Punti et al. 2009). Moreover, because of life stage, size or conditions, larvae specimens often cannot be identified to species level, leading sometimes to errors and imprecision in habitat and water quality (Sweeney et al. 2011).

MOLECULAR ECOLOGY

According to Weiss (1950), molecular ecology represents the continuum of biological interactions between cellular, molecular and organismal levels with the environment. However, as Lambert (1995) argued, Weiss definition was more about ecology itself. Molecular ecology is defined as biological interactions between cellular, molecular and organismal levels with the environment. These interactions are observed and studied by scientists thanks to the use of a wide array of DNA tools which allow them to map and explore interactions.

Molecular ecology is considered a recent field obtained by inter and multi-disciplinary cooperation amongst various fields. Molecular ecology is a branch of ecology, which has its main root in natural history, zoology, geography, geology, evolutionary studies, physics, bioinformatics until it reached molecular biology and populations genetic allowing ecology to branch into molecular ecology (Levin, 2010).

One of the major changes that molecular ecology has brought is molecular barcoding thanks

to the use of molecular markers. Molecular markers are sections of an organismal genome obtained by the polymerase chain reaction (PCR) (Kirsten et al. 2010). There are several types of molecular markers such as microsatellites, minisatellites, restriction fragment length polymerase and DNA sequence data. Molecular markers used in barcoding techniques, allow scientists to quantify genetic diversity, track individual's movement, measure inbreeding, characterise new species, identify species from mixed samples and retrace historical pattern of dispersal (Kirk et al. 2011).

Although markers have extremely changed and helped molecular ecologist, many studies are still limited to the narrow region of the genome, making it difficult to generalise about organisms and their evolutionary history (Narum et al. 2013). DNA-barcoding allows generating vast DNA libraries in order to identify unknown specimens through the use of standardise species-specific genomic regions called DNA barcodes (Shokralla et al. 2012).

The cytochrome c oxidase subunit 1 (CO1) gene region, is the mainly used to identify species across Animalia (Hebert et al. 2003); similarly, 16S ribosomal RNA (16S) is commonly used for bacterial identification (Sogin et al. 2006); the internal transcribed spacer (ITS) region is employed in fungi identification (Nilsson et al. 2008); finally, for plant identification, plastid DNA including maturase K and rbcl are used (Burgess et al. 2011). In recent years, DNA barcoding has gained popularity as a method of taxonomical identification of unknown species. However, for the past 250 years, taxonomy has been defined based on morphological structures such as the phenotypic characteristic of individuals (Friedheim, 2016). Morphological taxonomy has been the basis of all phylogenetic relationship hypothesis of extinct organisms based on fossil records.

Although Friedheim (2016) argue that morphological identification is been the basis of identification and it is more accurate, Galan et al. (2012), argue that there are too few taxonomic experts available for the many research disciplines. He also argued that morphological identification is complicated due to the different life-stages of some animals. larval and/or immature stages can be morphologically different from the adult phase, exactly like sexual dimorphism and cryptic and sister-species are often nearly-identical.

By the use of DNA-barcode, species can be taxonomically identified in a very short amount of time through the use of molecular techniques, is a non-invasive technique and is available to everyone and not just to expert (Dalziel et al. 2009). By sequencing an informative segment of DNA it is possible to define "molecular operational taxonomic units" (MOTU). in DNA barcoding literature, the designation MOTU has been widely used to describe clusters of sequences that act as representatives of the genomes from which they are derived generated by an explicit algorithm. Using a clustering algorithm, MOTUs can be defined by different approaches among which the use of specific cut-off values based on sequences similarity. In DNA barcoding literature MOTU can designate different situations that usually are divided into three distinct groupings: a group of unidentified organisms sharing similar sequences; a group of organisms within a species that are distinct at the molecular level from other members of the species; a group of organisms from different species that are similar at the molecular level (Galimberti et al., 2012).

One of the reasons for molecular ecology rapid advancement as a field of study is the polymerase chain reaction. PCR allows the amplification of billions of copies of a specific piece of DNA from the genome with just a few starting copies. Moreover, it is now possible to isolate DNA from hair, urine, shed skin and faeces. This was a massive innovation for the molecular field because PCR is considered a non-invasive technique by which is possible to obtain massive results with just a small starting amount of DNA avoiding ethical issues of killing organisms just for study purposes (Kirsten et al. 2010). Molecular ecology techniques would probably be not as useful as they are without bioinformatics analysis.

Bioinformatics is defined as the application of tools of computation and analysis to the interpretation of biological data and it is an interdisciplinary field fundamental for molecular biology and molecular ecology (Bayat, 2002; Excoffier et al, 2005). The use of computers for molecular ecologists, starts in the laboratory. For instance, when a DNA molecule has to be cut and tailored with one or two of the hundreds of enzymes reagents available, millions of possibilities have to be analysed. With the use of a simple computer software, all the possible fragments can be analysed, and

the right combination could be suggested in a very short amount of time (Pongor et al. 1999).

The main tools used in bioinformatics are computer software programs and internet. Anyone with access to the internet and websites can now use basic bioinformatics tools although this does not imply that everyone can analyse easily genomic data and other tools (Bayat, 2002). Bioinformatics is now being used for various important task in the ecological and molecular field, including analysis of gene variation and expression, gene and protein structure prediction and function, sequences analysis and presentation and analysis of molecular pathways in order to understand and study gene-disease interactions (Bayat, 2002).

Probably the biggest bioinformatics tool used by molecular ecologist are databases. Since various species genome have been sequenced after the genome process, a necessitation of computer databases that feature rapid assimilation, easy format and algorithm were required. Nowadays, a number of databases containing useful information are available online and easy to access. One of the simplest and better-known search tools is the basic local alignment search tool (BLAST). This algorithm software allows users to search other databases for genes with similar nucleotide structure. Moreover, it allows comparisons of unknown DNA with millions of sequences from all kind of organisms until a match is found (Bayat, 2002).

DNA BARCODING AS A MOLECULAR TOOL

The term ‘DNA barcoding’ is of recent use in the literature. It relies on the use of a standardized-DNA-region as a tag for rapid and accurate species identification (Valentini et al. 2009). The aim of DNA taxonomy is to find molecular-defined operational taxonomic units (MOTU) (Floyd et al. 2002) on the basis of sequence differences at short, orthologous marker gene sequences (Tautz et al, 2003).

The major goal of DNA barcoding is identification of species. Morphological species-level

identification is possible for males and some later instars of aquatic invertebrates, but currently is not feasible for females and immatures of many arthropod groups (and many other phyla).

A second goal of DNA barcoding is the diagnosis of new species. In practice, new species are diagnosed when specimens of a series are sequenced and the specimens demonstrate genetic differentiation that exceeds the barcode gap being used at the time. DNA barcoding relies on seven different steps: (Valentini et al, 2009) sampling in the field; DNA extraction, DNA amplification with universal primers; High throughput parallel pyrosequencing; reference database; species identification via DNA barcoding and biodiversity description.

The cost of a qualified taxonomist is substantial, and, in some cases, taxonomic expertise may not be available. In this situation, researchers hope to assess the diversity in large samples through molecular means. Sequenced specimens are compared to sequences in an existing library to determine if exact or close matches exist in the library (Hebert et al. 2016). Specimens outside the barcode gap threshold used for the study are treated as belonging to a MOTU. Even if morphological identification of a species is possible, DNA barcoding might enhance biodiversity inventories by being faster and cheaper, and by overcoming the taxonomic impediment. It could allow biodiversity assessment through the identification of taxa from the traces of DNA present in environmental samples such as soil or water (Kranzfelder et al. 2008). The use of DNA-barcoding will not be necessary for assessing the biodiversity of well-known ecosystems. However, in ecosystems showing high species richness, such as those in tropical environments, it is unrealistic, within a limited time period, to identify all animals and plants by morphology alone.

The ideal DNA barcoding system should meet the following criteria (Kranzfelder et al. 2008): The gene region sequenced should be nearly identical among individuals of the same species, but different between species; it should be standardized, with the same DNA region used for different taxonomic groups; the target DNA region should contain enough phylogenetic information to easily assign unknown or not yet ‘barcoded’ species to their taxonomic group (genus, family, etc.); it should be extremely robust, with highly conserved priming sites and highly reliable DNA amplifications and

sequencing. This is particularly important when using environmental samples, where each extract contains a mixture of many species to be identified at the same time. Finally, the target DNA region should be short enough to allow amplification of degraded DNA. Usually, DNA regions longer than 150 bp are difficult to amplify from degraded DNA

CYTOCHROME OXIDASE SUBUNIT 1 (CO1)

The mitochondrial protein-coding gene, cytochrome c oxidase subunit 1 (CO1), is a widely accepted marker for molecular identification to the species level across diverse taxa (Hebert et al. 2004). CO1 meets all of the DNA Barcoding System criteria listed above and has been accepted as a practical, standardised, species-level DNA barcode for many groups of animals, but not in plants and fungi (Kress et al. 2015). According to Goodwal-Copestake et al. (2012), CO1 popularity as DNA-barcode derives from the availability of several sets of conserved PCR primers (such as Folmer et al. 1994), and from its dual purpose for estimating both intra-specific variation and inter-species identification.

AIMS AND OBJECTIVES

In this project, the community structure and diversity of a group of freshwater invertebrates (non-biting midge larvae) has been studied in relation to heavy metal pollution and other chemical and magnetic properties of sediments and water from ponds in Wales. The main purpose was to assess the effect of pollutants and environmental conditions of sediments and water on invertebrate biodiversity while comparing suspected polluted and non-polluted ponds in Wales.

Although environmental pollution can be considered natural, for example erupting volcano spews out huge quantities of rocks, ash, chlorine, sulfur dioxide, and other chemicals, the work presented here emphasizes on the effect of anthropogenic pollutants (i.e., pollutants produced by human activity) on biodiversity.

The aims of this study were:

- To characterise the levels of heavy metals pollution in ponds and their possible influence on the chironomids.
- To identify non-biting midge (chironomids) larvae in ponds in Wales through DNA barcoding.
- To characterise the genetic diversity and genetic structure of non-biting midges among ponds in Wales.
- To characterise the community structure of non-biting midges in Wales.

It was hypothesised that the ponds studied in Wales would have different levels of environmental conditions and magnetic susceptibility, and that these conditions would also be reflected in the DNA-barcodes of non-biting midges. It was expected to find higher diversity and different community of chironomid larvae in non-polluted ponds than in polluted ponds.

Materials and methods

SAMPLING SITES

The sampling of sediment, water and chironomid larvae took place between the 27th and the 30th of May 2017 and between the 25th and 30th of May 2019. All the chironomid larvae and sediments samples were collected from freshwater ponds in Parys Mountain (Anglessey, Wales) and Cwm-Idwal

(Snowdonia National Park, Wales) (Fig. 2). The two main sampling sites (Parys Mountain and Cwm-Idwal) were located about 48 km apart from each other (coordinates are found in table S1). These sampled regions were chosen in order to have samples from suspected polluted ponds (Parys Mountain mine) and samples from suspected non-polluted ponds (Snowdonia National Park). A total of seven ponds were studied, including five ponds from Parys Mountain (namely, WAPM1, WAPM2, WAPM3, WAPM4, WAPM5) and two from Cwm-Idwal (namely, WACI1 and WACI2) (Fig. S 1). One pond from Parys Mountain was suspected to be non-polluted (WAPM1) because it has not been used as part of the desiccation ponds for the process of metal extraction, it is a slightly higher elevation than the other ponds in Parys Mountain (20 m of elevation compared to the suspected polluted ponds) and water seems to be only obtained through rainfall and runoff from adjacent hills (personal information). Also, the two ponds from Cwm-Idwal were suspected to be non-polluted as there have been little human activities in the area, and the water in these ponds derives from mountain streams (Tonkin et al. 2014).

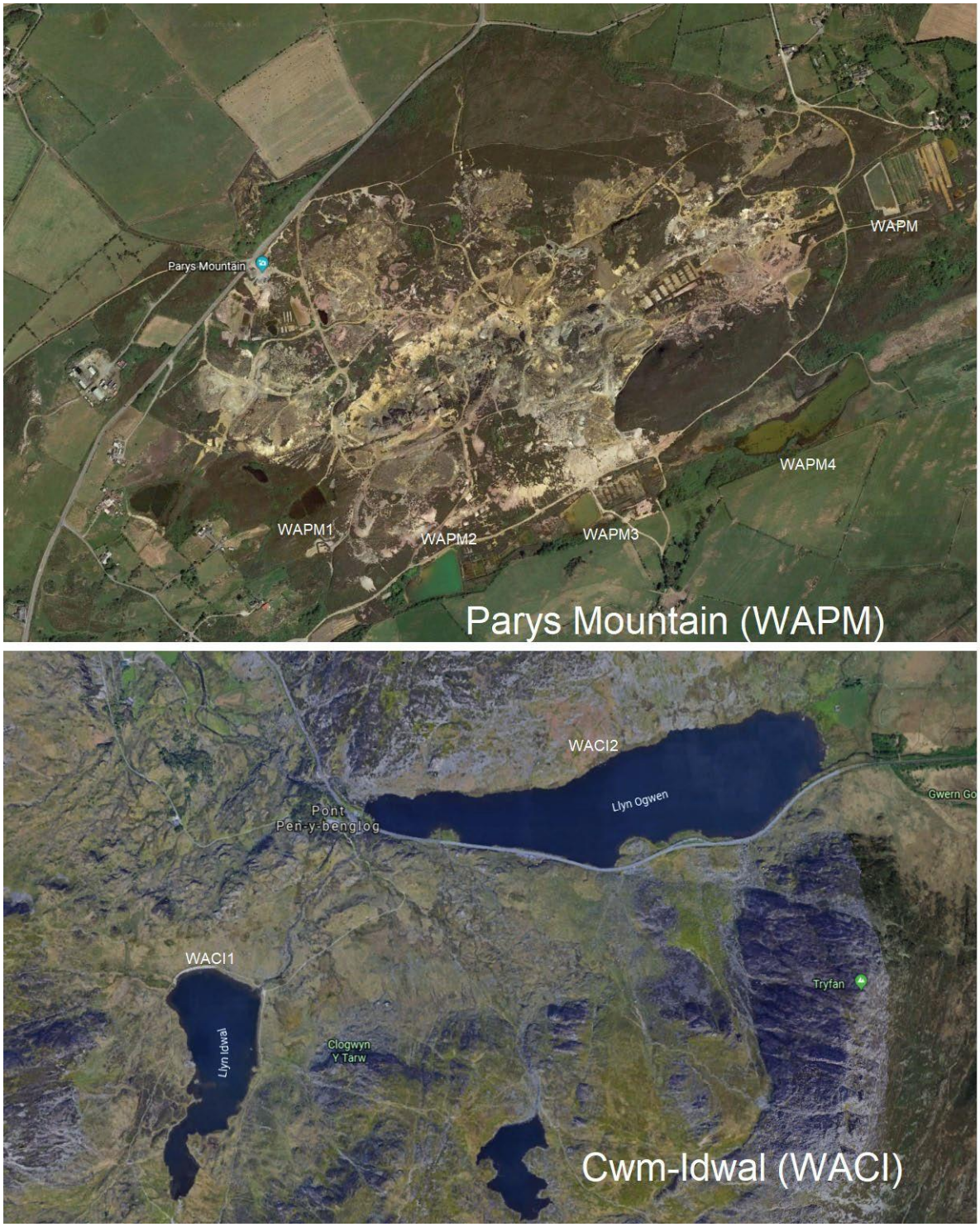


Figure 2. Map of the sampling area.

The geomorphology of Cwm-Idwal is characterised by a moraine–mound complex ('hummocky moraine') located on both the east and west of Llyn [lake] Idwal. These moraines have been the subject of numerous investigations (e.g. Darwin, 1842; Escritt, 1971) due to their importance for

understanding the significance of Younger Dryas glaciers in the British Uplands. The majority of the moraines are 8 to 80 m in length, with the exception being a set of discontinuous stream-breached ridges totalling about 450 m in length which are stacked against the western cirque wall (Tonkin et al. 2014). The rocks that form Parys Mountain dated to the Ordovician and Silurian age.

Mineralisation occurred in several phases: sediments were being laid down associated with volcanic activity on the sea bed which then lead to a localized concentration of sulphide minerals which then folded into a vertical inclination (Vernon 1996). During later phases, some of the minerals were remobilised and impregnated in the surrounding rocks (Vernon, 1996)

CHIRONOMID LARVAE, SEDIMENT AND WATER SAMPLING

Chironomid larvae and sediments samples were collected from three different sites on each pond to obtain a representative sample from each pond (Fig. S1). Pond sediments were collected in 50 mL Falcon tubes and labelled according to the site of collection.

Freshwater invertebrates were collected by kick-sampling for approximately three minutes by kick-sampling using a standard 1 mm mesh pond (Traister et al. 2013). Chironomid larvae were then visually separated from other insect larvae in the field and stored in labelled 2.0 mL screw cap plastic tubes containing 96% ethanol to preserve them.

SEDIMENT ANALYSES

Salinity, pH, temperature, total dissolved solids and dissolved oxygen were measured at each site within each pond by using a multi-parameter water quality meter (Hanna HI9143 for dissolved oxygen levels, HI9835 for electric conductivity, salinity, temperature and total dissolved solids and

Jenway 570 for pH). Three water samples per pond were collected and stored in 50 mL Falcon tubes and 2 ml of HNO_3 (65%) were poured into the water samples to stabilize it.

For estimating the magnetic susceptibility, a total of 21 sediment samples were collected in 50 mL Falcon tubes (three per pond). All the sediments were filtered and dried in a Memmert heater and drying oven for 24 hours at 50° C. The dried samples were than subsampled and packed into 10 cm^3 plastic sample pots wrapped with clean films to keep them tight. MS was measured using a Bartington instrument MD2B susceptibility meter. For sediments samples, a low-field magnetic susceptibility was the best parameter to be tested because of the small grain-size (Chernicoff, 1984).

The isothermal remanent magnetization (IRM) of all the 21 samples was measured using the same 10 cm^3 plastic sample pots. An ASC Scientific IM-10 impulse magnetiser was used to impart an IRM in an applied field of 1000 mT (or 1 T). A Molspin Minispin spinner magnetometer was used to measure the IRM.

(Robertson, D. et al. 1993). For sediments samples, a low-field magnetic susceptibility was the best parameter to be tested because of the small grain-size (Chernicoff, 1984).

The isothermal remanent magnetization (IRM) of all the 21 samples was calculated. A ASC Scientific IM-10 impulse magnetizer was used to impart the IRMs. The samples were then subjected to a progressively increasing unidirectional magnetic field (+1000, -100, -300, -1000 mT) according to the sample's attenuation. After each increase in applied field the sample were removed and their magnetisation (i.e. IRM) were measured using a Molspin Minispin spinner magnetometer (Robertson, D. et al. 1993).

The presence of metals (Iron, Fe; Zinc, Zn; Lead, Pb and Copper, Cu) was detected by Inductively Coupled Plasma-Optical Emission Spectroscopy (ICP-OES). ICP-OES is used in atomic spectroscopy, and during analysis the sample is decomposed by intense heat into a cloud of hot gases containing free atoms and ions of the element(s) of interest (Fitzsimmons, 2015). The high temperatures cause significant amounts of collisional excitation and ionization of the sample atoms. Once the atoms or ions are in their excited state, they can decay to lower states through thermal or

radiative (emission) energy transitions. During ICP-OES analysis the intensity ofthe light emitted at specific wavelengths is measures and used to determine the concentration of the element(s) of interest. In ICP-OES analysis the thermal excitation sources can populate a large number of different energy levels for several different elements at the same time (Fitzsimmons, 2015). All of the excited atoms and ions can then emit their characteristic radiation at the same time. This results in the flexibility to choose from several different emissions concurrently and allows detection of multiple elements concurrently.

All 21 sediments samples digested according to the protocol (Table 1) using the Speedwave Four Berghof Microwave Digestion.

Table 1. Protocol used in the Microwave Digester.

Step	T (°C)	P (bar)	Ta (min)	Time (min)	Power (%)
1	180	30	2	25	90
2	50	25	1	10	0
3					
4					
5					

The same material that was studied magnetically, was weighted between 0.5-1.0 g and 2.5 mL of HNO_3 (65%) and 7.5 mL of HCl (37%) were added into the vessel before the digestion. The digested samples were stored in 15 mL Falcon tubes at room temperature. The water samples collected form the ponds were filtered and stored HNO_3 (65%) acid at 8°C. A stock solution was prepared containing: 20 mL (1000 ppm) of Fe and Zn; 5 mL (1000 ppm) of Cu and Pb; 2 mL of HNO_3 65% and diluted with distilled water to a final volume of 100 ml. Four standards for the calibration curve were made at different concentrations (Cu and Pb: 0.0, 1.25, 2.5 and 5 ppm; Fe and Zn: 0, 5, 10 and 20 ppm) by adding respectively 0, 2.5, 5 and 10 mL from the stock solution and diluting up to 50 mL with distilled water.

The digested sediments samples, the filtered water samples and the four standards were run into an ICP-OES Optima 8000. The sediments sample's weight data were inserted into the ICP-OES protocol for standardisation of the data. The samples that were not within the calibration curve were re-diluted 100 times and they were re-run. The sediments data obtained from the ICP-OES were then converted from mg/L into mg/Kg.

STATISTICAL ANALYSIS

As the 21 samples (3 replicates for each pond) resulted non-normally distributes, a Kruskall-Wallis test was performed to compare differences in: the environmental conditions (pH, temperature, electric conductivity, total dissolved solids, salinity, dissolved oxygen). MS and IRM were also tested for differences using Mann-Whitney U test across the 3 replicates of the seven ponds.

Principal component analysis (PCA) was performed across all the replicates of the 7 ponds in order to reduce the dimensionality of the dataset (pH, temperature, electric conductivity, total dissolved solids, salinity, dissolved oxygen, magnetic susceptibility and isothermal remanent magnetization). Discriminant analysis (DA) was tested on the environmental factors (pH, temperature, electric conductivity, total dissolved solids, salinity, dissolved oxygen) across all the replicates of the 7 ponds. PCA is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables (Abdi et al. 2010). Its goal is to extract the important information from the table, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as clusters (Abdi et al. 2010). DA, is a statistical method which aims is to discriminate variables between two or more naturally occurring groups (Koklu et al. 2010). It calculates mathematical weights for scores on each discriminator variable that reflect the degree to which scores on that variable differ among the groups being discriminated. It forms one or more

weighted linear combinations of discriminator variables called discriminant functions (Koklu et al. 2010). PCA and DA were carried out using Minitab 19, (Minitab, LLC 2019).

MOLECULAR ANALYSIS

A total of 120 Chironomid larvae were dissected under a dissection microscope by cutting the head capsule and the last part of the abdominal segment for the future morphological identification and comparison with DNA barcoding results. The dissected parts were preserved in 75% ethanol at -20°C. The intact part of the abdomen part was then used for DNA barcoding. Total genomic DNA was extracted from only 78 Chironomidae larvae as the remaining didn't amplify correctly, using the GeneJET genomic DNA purification kit (ThermoScientific)nfollowing the protocol but with a few modifications: 20 µL of RNase A solution were added, and the samples were centrifuged at a higher speed (8000 rpm).

A fragment of 680 base pairs (bp) of the mitochondrial gene Cytochrome c oxidase subunit 1 (CO1) was amplified using the primers: LCO1490 (5'-ggtcaacaaatcataaagatattgg-3') and HC02198 (5'-taaacttcagggtgaccaaaaaatca-3') (Folmer et al. 1994). PCR was performed in 50 µL total volume including: 25 µl Dream*Taq* Green PCR Master Mix (ThermoFisher), 2.5 µL of each primer (forward and reverse, at 10 µM), 18 µL water and 2 µL DNA. A negative control was made for each PCR containing 2 µl water instead of DNA. PCR cycling conditions were as follows: an initial step of 95°C for 5 min, followed by 35 cycles of 95° C for 1 min, 40°C for 1 min and 72°C for 1.5 min, and then a final extension of 72°C. PCR products were examined using 1% agarose gel electrophoresis stained with SYBR safe (Invitrogen) and visualised under UV light using a Bio Rad Gel Doc XR+ system. Successful PCR products were purified using GenJET purification kit (ThermoFisher)

following the protocol but with a few modifications: all the centrifugations were carried out at 32000 rpm and only 30 µL of elution buffer were added and incubated at room temperature for 5 minutes

instead of 1; and examined again using 1% agarose gel electrophoresis. DNA sequencing was performed by DBS Genomics (Durham University, UK). The DNA sequences were visualized and manually edited using BioEdit v 7.0.5.3 (Hall 1998).

The DNA sequences were then identified and assigned a "putative" identification name (either to species, genus, or any other taxonomic level) based on the first hit by BLAST (NCBI). In order to have a high level of confidence, putative genus name >97% match were considered a good match with the GenBank samples (Altschul et al.1990). These Operational Taxonomic Units (OTUs), were then used for downstream bioinformatic analysis. The max score, total score, % ID, e-value and query were also retrieved from BLAST. Where the max score is defined as the highest alignment score of a set of aligned segments from the same subject sequence; the total score is defined as the sum of alignment scores of all segments from the same subject sequence; The percent identity is a number that describes how similar the query sequence is to the target sequence; the e-value is defined as the default sorting metric and query is defined as the percent of the query length which is included in the aligned segments (NCBI, 2019).

PHYLOGENETIC ANALYSIS

A total of 78 sequenced individuals from Wales were obtained, and 539 CO1 sequences were retrieved from GenBank. For each species, duplicates were removed in order to obtain a clearer view of the clustering of the Wales' species in the phylogenetic tree. Sequences were aligned using Clustal-W multiple sequence alignment in MEGA version 6 (Thompson et al. 1994*).* A Neighbour-Joining tree (NJ) was generated in MEGA 6 using Kimura-2-Parameter (K2P) model, 1000 bootstrap replicates and *Aedes albopicutus* MK736660.1 (Diptera; Culidae) as outgroup (Wilkerson et al., 2015). The tree was then modified by using FigTree version 1.4.4 software (Rambaut, 2007).

ASSESMENT OF GENETIC AND COMMUNITY STRUCTURE

DNA polymorphism was analysed for all the 78 samples using DnaSP version 5.10.1 (Librado et al. 2009), estimating the number of haplotypes (h) and haplotype diversity (Hd) and nucleotide diversity (π) per pond.

Genetic differentiation among the seven ponds was estimated using Wright's fixation index (F_{ST}) (Wright 1951). F_{ST} e can provide the basis for a measure of genetic distance when divergence is caused by drift (Reynolds et al. 1983). The Fst value can be low (> 0.05), moderate (between 0.05 and 0.15) or high (>0.25) (Balloux et al. 2002). A phylogenetic haplotype network was built by using Network 4.5.1.6 Software (Fluxus Technology, 1999). Phylogenetic networks are connected graphs with cycles. Haplotype or allele networks are generated by, for example, a median-joining (Bandelt et al. 1999) or statistical parsimony (Clement et al. 2000) analysis, in which nodes represent different allelic sequences, joined by edges (branches) whose length is defined and shows the number of nucleotides that differ between them (Mardulyn, 2005).

COMMUNITY STRUCTURE

The analysis on community structure was performed by using PRIMER version 6 software using the presence/absence data and the relative abundance of putative genera of chironomids. The software PAST v14 (Hammer et al. 2001) was used in order to calculate the chironomid genus richness and the genus diversity index. In this research, the genus richness was calculated by counting the numbers of putative genera identified in total and at each site. The genus diversity index for Parys Mountain and Cwm-Idwal and per site was obtained using Simpson's diversity index (D) as implemented in

PAST : $D = 1 - \sum \left(\frac{n}{N}\right)^2$

A Bray-Curtis dissimilarity among sites and cluster (dendrogram) analysis was performed

using Minitab version 19. The dendrogram provides a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data (Shrestha and Kazama, 2007).

A multi-scale-dimensional plot (MSD) was created to look at the differences in community structure among the seven ponds, using Bray-Curtis similarity index. Bray Curtis similarity index was chosen as resemblance analysis for the multiscale dimensional plot (MDS) although no overall transformation or weighting has been applied to the data. According to Clarke et al. (2006), there is no transformation, on the normal functioning of a Bray–Curtis analysis when at least a modest amount of data is present for all samples. According to Hout et al. (2013), *stress function* was calculated. Lower stress values indicate a better fit; thus, the algorithms attempt to increase the fidelity to the input data by minimizing this stress function (Hout et al. 2013).

Moreover, to test for differences in species diversity, a non-parametric Analysis of Similarities (ANOSIM) was performed from the Bray-Curtis similarity matrix (Marchant et al. 2000). ANOSIM is based on a nonparametric permutation procedure applied to the rank similarity matrix (in this case, Bray–Curtis similarity) that compares the degree of separation between predefined groups with the test statistic, R (Clarke and Warwick 2001). Values of R near 0 indicate no distinguishable separation between groups, whereas values near 1 indicate complete separation. The R test statistic is first calculated as a global test to determine if differences are present between groups (Zuellig and Schimdt, 2012).

Results

ENVIRONMENTAL CONDITIONS AND SEDIMENT ANALYSIS

Kruskal-Wallis test was conducted on the environmental conditions: pH, temperature, salinity, conductivity, dissolved solids and dissolved oxygen shown in figure 3. Three replicates (A,B and C)

were analysed for each of the 7 ponds. The results indicated a significant difference between suspected non-polluted ponds (WAPM1,WACI1,WACI2) and most of the suspected polluted ponds (WAPM2,WAPM3,WAPM4,WAPM5).

Specifically, Results from the Kruskal-Wallis test performed on pH showed a significant difference amongst the 7 ponds ($p= 0.004$). A significant difference ($P<0.05$) was found between the WAPM1 and WAPM2,3,4 and WAPM5. Likewise a significant difference was obtained between WACI1 and WAPM4 and 3 also between WACI2 and WAPM4.

Results from the Kruskal-Wallis test performed on temperature and dissolved oxygen showed no significant difference amongst the 7 ponds ($p> 0.05$).

Results from the Kruskal-Wallis test performed on salinity showed a significant difference amongst the 7 ponds ($p= 0.00$). A significant difference ($P<0.05$) was found between WAPM1 and WAPM2,3, 4 and WAPM5; between WACI1 and WAPM2,3, 4 and WAPM5; between WACI2 and WAPM2,3, 4 and WAPM5.

Results from the Kruskal-Wallis test performed on conductivity showed a significant difference amongst the 7 ponds ($p = 0.003$). A significant difference ($P<0.05$) was found between WAPM5 and ponds WAPM1,2,3, 4, WACI1 and WACI2.

Results from the Kruskal-Wallis test performed on dissolved solids showed a significant difference amongst the 7 ponds ($p < 0.05$). A significant difference ($P<0.05$) was found WAPM1 and WAPM3,4 and WAPM5; between WACI1 and WAPM2,3, 4 and WAPM5; between WACI2 and WAPM2,3, 4 and WAPM5.

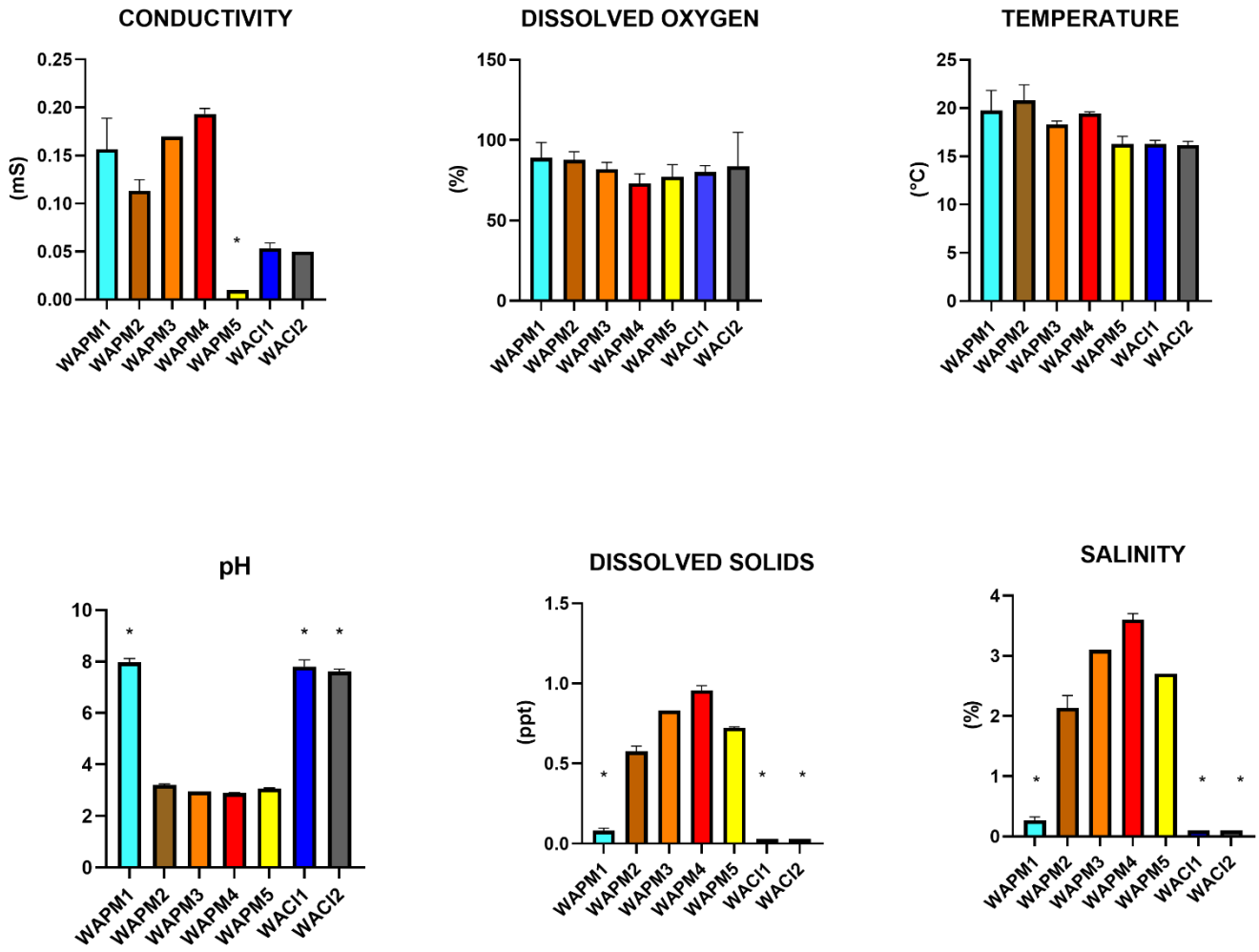


Figure 3. Kruskal- Wallis test on the six environmental conditions. *= significantly different.

There were no differences in magnetic susceptibility (MS) and Isothermal remanent magnetization (IRM) among the 7 ponds (p=0.206; p=0.653), (Table 2). However, WAPM2 and WAPM3 showed high peak in comparison to all the other sites (Fig. 4).

Table 2. Mann-Whitney-U test results on magnetic susceptibility (MS) and isothermal remanent magnetisation (IRM).

	p-value	Confidence level (%)
MS	0.206	95.45
IRM	0.653	95.45

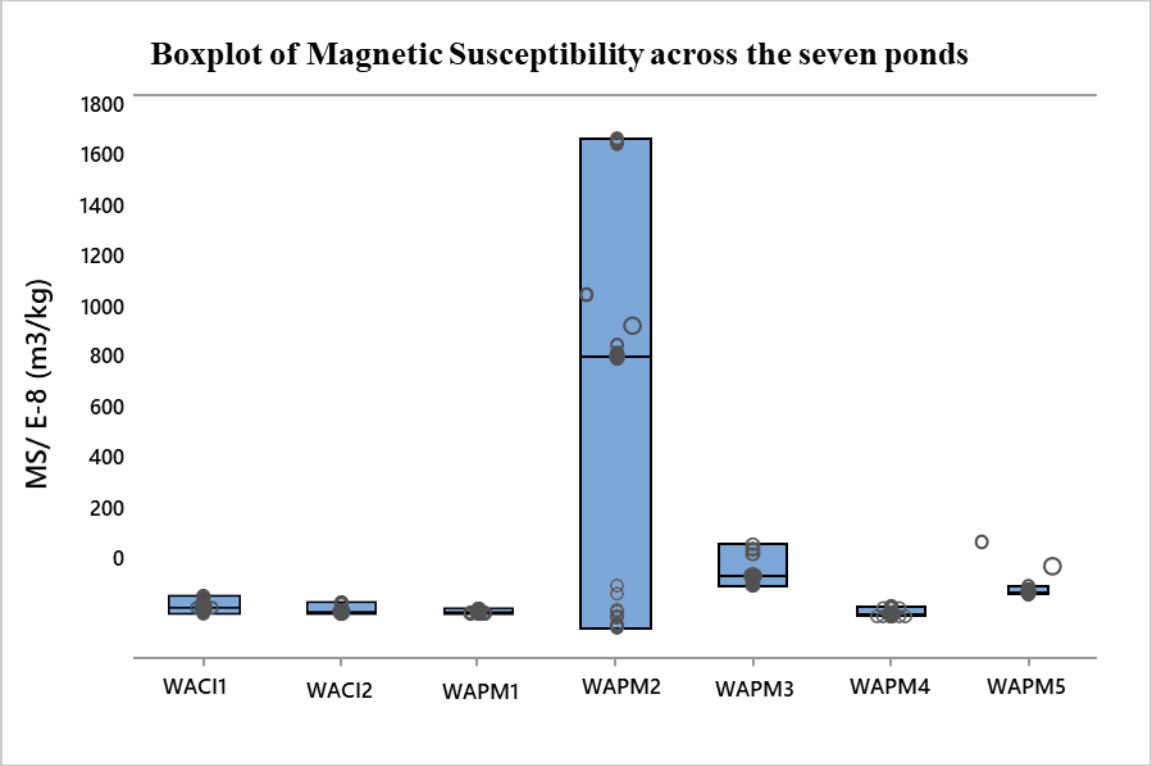


Figure 4. Magnetic susceptibility among the seven ponds in Wales. Although differences were observed, these were not statistically significant (MW p=0.20 >0.05).

MAGNETIC PROPERTIES

The linear regressions of the calibration curves showed a high correlation coefficient for all heavy metals (Zn = 0.999979, Pb = 0.999961, Cu= 0.999825 and Fe= 0.999944), which allowed the prediction of heavy metal concentration in sediment and water samples (see appendix Fig. S2). Sediments sample data obtained after the dilutions calculations (Table 3) showed significantly higher values of metals composition in the suspected polluted ponds (WAPM2, WAPM3, WAPM4 and WAPM5) than in the suspected non-polluted ones. Cu and Fe resulted as the highly concentrated metals in the suspected polluted ponds and especially in pond WAPM3, WAPM4 and WAPM5. Moreover, across the suspected non-polluted ponds, pond WAPM1 showed slightly higher level of metals concentration in comparison to the suspected non-polluted ponds in Cwm-Idwal.

Table 3. Concentrations in calibration units of Cu, Fe, Pb and Zn sediment samples from Wales.

Sample Id	Cu (mg/Kg)	Fe (mg/Kg)	Pb (mg/Kg)	Zn (mg/Kg)
P1A	232,78	180,58	280,07	565,78
P1B	85,41	177,04	399,87	300,56
P1C	77,57	194,17	321,93	1335,95
P2A	2307,70	629,50	4059,86	2128,66
P2B	2324,19	34972,28	5554,52	6219,79
P3A	66,95	33685,57	8,01	48,78
P3B	1222,99	124960,52	12,47	34,46
P3C	1403,38	32017,97	190,56	31,73
P4A	665,85	634670,71	219,21	208,07
P4B	201,56	333611,23	139,18	73,98
P4C	117,42	200367,74	28,97	55,26
P5A	2986,78	334201,02	20464,84	3577,60
P5B	1715,81	105633,83	14038,50	647,14
C1A	12,95	178,79	47,46	57,27
C1B	11,18	171,98	14,10	44,60
C1C	13,38	178,79	13,50	70,62
C2A	10,73	167,20	45,83	42,69
C2B	17,08	188,10	36,20	47,54
C2C	15,47	207,56	64,84	36,84

Water samples (Table 4) data showed similar results to the sediments data as the suspected polluted ponds showed higher concentrations of metals than the suspected non-polluted ones.

Table 4 Concentrations in calibration units of Cu, Fe, Pb and Zn water samples from Wales.

Site	Cu (mg/L)	Fe (mg/L)	Pb (mg/L)	Zn (mg/L)
P1AW	0.075	1.686	0.034	0.066
P2AW	3.77	4.115	1.373	6.368
P2BW	3.925	3.522	1.383	6.637
P2CW	3.677	3.283	1.321	6.234
P3AW	3.21	10.253	0.563	8.946
P3BW	0.3157	10.43	0.564	9.086
P3CW	3.221	10.698	0.578	9.333
P4AW	2.518	4.468	0.269	11.039
P4BW	3.218	15.454	0.275	12.947
P4CW	2.99	15.822	0.26	12.081
P5AW	6.911	1.305	5.662	9.973
P5BW	10.797	3.419	1.302	18.932
C1AW	0.12	0.164	-0.007	0.161
C1BW	0.056	0.288	-0.014	0.092
C1CW	0.058	0.142	-0.012	0.058
C2AW	0.048	0.116	-0.016	0.067
C2BW	0.06	0.123	-0.013	0.092
C2CW	0.049	0.119	-0.015	0.146

PCA of the 16 parameters (see appendix, Table S2) showed that four PCs explained about 81.7% of the total variance across the seven ponds in Wales. According to the eigenvalue-one criterion, just the eigenvalues >1 were considered essential (Fig. 5).

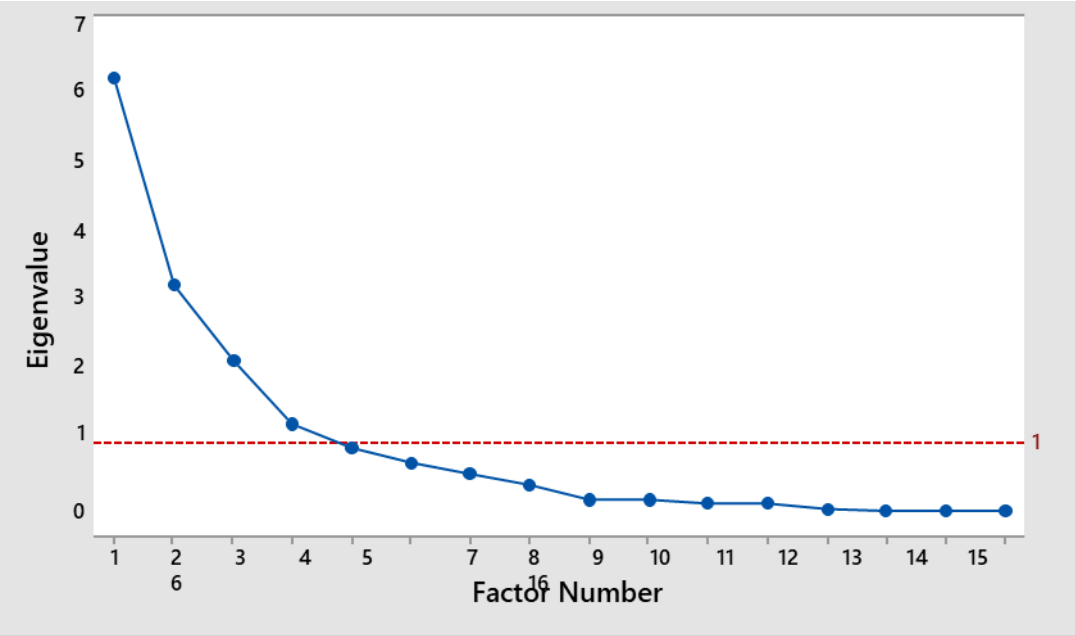


Figure 5. Scree plot of the Eigenvalue across the sixteen components.

In these results, the first four principal components had eigenvalues greater than 1. The scree plot (Fig. 5) showed that the eigenvalues start to form a straight line after the third principal component. The first four principal components were chosen for further analysis as 99.2% was considered an adequate amount of variation explained in the data.

The first principal component accounts for 41.7% of the total variance (see appendix, Table S2). The variables that correlate the most with the first principal component (PC1) are Total dissolved solids (0.337), Salinity (0.337), Zinc concentration in the water (0.338), and Copper present in the water (0.321), (Table 5). The first principal component is positively correlated with all four of these variables. Therefore, increasing values of these for variables increase the value of the first principal component. The second principal component accounts for 61.1% of the total variance. The variable that correlates the most with the second principal components (PC2) were electric conductivity (0.416) and Fe concentration in the water (0.380).

Table 5. Eigenvectors of the principal component analysis. Only values above 0.3 were taken, only first 4 components.

Eigenvectors				
	PC1	**PC2**	**PC3**	**PC4**
Variables				
pH	-0.349	-0.106	-0.048	0.178
TEMPERATURE (°C)	0.117	0.13	0.507	0.217
ELECTRIC CONDUCTIVITY (mS	0.085	0.416	0.252	0.313
TOTAL DISSOLVED SOLIDS (ppt)	0.337	0.206	-0.021	-0.058
SALINITY (%)	0.337	0.206	-0.024	-0.055
DISSOLVED OXYGEN (%)	-0.121	-0.083	0.416	-0.07
MS/ E-8 m3/kg	0.081	-0.009	0.326	-0.583
IRM1T / Am2/kg	0.11	-0.185	0.428	0.078
Copper	0.27	-0.301	0.079	0.167
Iron	0.215	0.15	-0.213	0.342
Lead	0.219	-0.385	-0.207	0.099
Zinc	0.159	-0.347	0.277	0.372
Water (Copper)	0.321	-0.138	-0.119	-0.107
Water (Iron)	0.226	0.38	0.005	0.054
Water (Lead)	0.243	-0.329	-0.076	0.052
Water (Zinc)	0.338	0.108	-0.134	-0.006
Simpson's index	-0.278	0.091	-0.076	0.395

Results of the PCA of the score plot for the first two components showed three clusters of samples based on their similarities (Fig. 6): the first cluster (C1), composed by the suspected non polluted ponds (WAPM1, WACI1, WACI2 and by one of the three ponds of WAPM5); the second cluster (C2), composed by two of the suspected polluted ponds (WAPM4 and the two ponds of WAPM5) and the last cluster (C3) composed by two of the suspected polluted ponds (WAPM3 and WAPM4). C1 and C2 were found to be different based on PC1, such differences are likely to be due to the factors that have heavy influences on PC1. Because of the similarity between the suspected polluted ponds and WAPM5, an outlier plot was made (see appendix, Fig. S3). As no point was found above the reference line, no outliers were recognized in the plot.

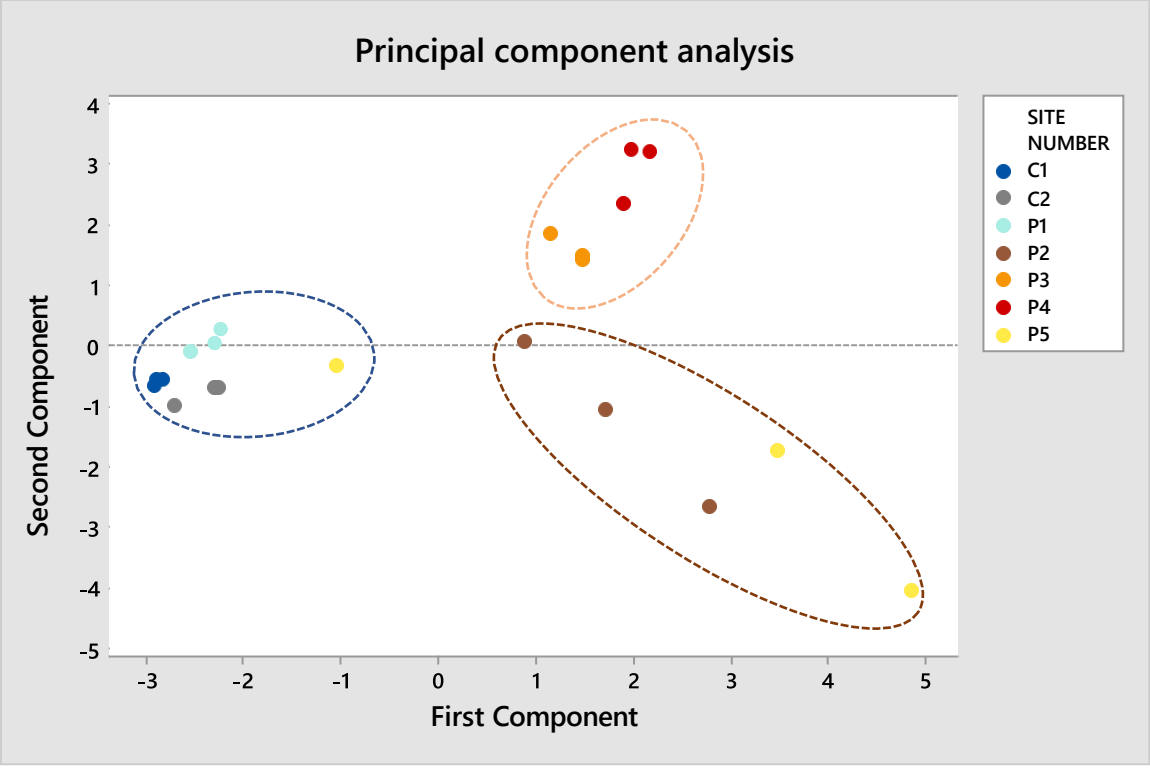


Figure 6. Score plot of the autoscalated data. Points are coloured according to the site number were C1 and C2 represent Cwm-Idwal ponds and P1,2,3,4 and 5 represent Parys Mountain.

Results for discriminant analysis showed that overall, 90.5% of observations were placed into the correct group (see appendix, Table S3). Ponds P1-P5 (WAPM1,2,3,4 and 5) had the highest proportion of correct placement, with 100% of the observations correctly placed. Ponds C1-C2 (WACI1 and WACI2) had the lowest proportion of correct placement, with only 2 of 3 observations, or 66.7%, correctly classified (see appendix, Table S4). Therefore, the classification system has the most problems when identifying observations that belong to ponds C1 and C2. Squared distance results (see appendix, Table S5) demonstrated that suspected non-polluted ponds (WAPM1, WACI1 and WACI2) have a relatively small squared distance and therefore they are classified more similar to each other and different from the suspected polluted ponds.

From The analysis, 19 out of 21 objects were classified into the correct place. The Summary of Misclassified Observations table (Table S6) shows observation 16 was misclassified into group C2 instead of its true group (C1) and that observation 19 was misclassified into group C1 instead of its true group (C2) which was the most frequent misclassification.

DNA SEQUENCES AND BLAST

In total, 120 DNA extractions were carried out. From these, 78 samples (WAPM1=13, WAPM2=14, WAPM3=15, WAPM4=10, WAPM5=4, WACI1=15 and WACI2=7) gave successful PCR amplifications and provided good quality DNA sequences for analysis (65% success). The 78 DNA sequences from Wales were compared against GenBank using BLAST, and the first hits were recorded. A total of 615 DNA sequences were aligned and analysed (78 from Wales and 537 from GenBank). All DNA sequences obtained had a minimum length of 600 base pairs (bp).

The putative identification of MOTU from Wales was carried out with BLAST. The sixteen different genera in the family Chironomidae and average GenBank identity from BLAST results by site and by genus are shown in Table 6. Some of the samples from Wales (23 mtDNA sequences) had

hits in GenBank with genus and species names recorded; however, the other 55 mtDNA sequences had hits in GenBank with only genus name recorded; therefore, all analyses in this study were done at the genus level rather than at the species level. In 16 mtDNA sequences, the first and second hits returned the same genus names in GenBank.

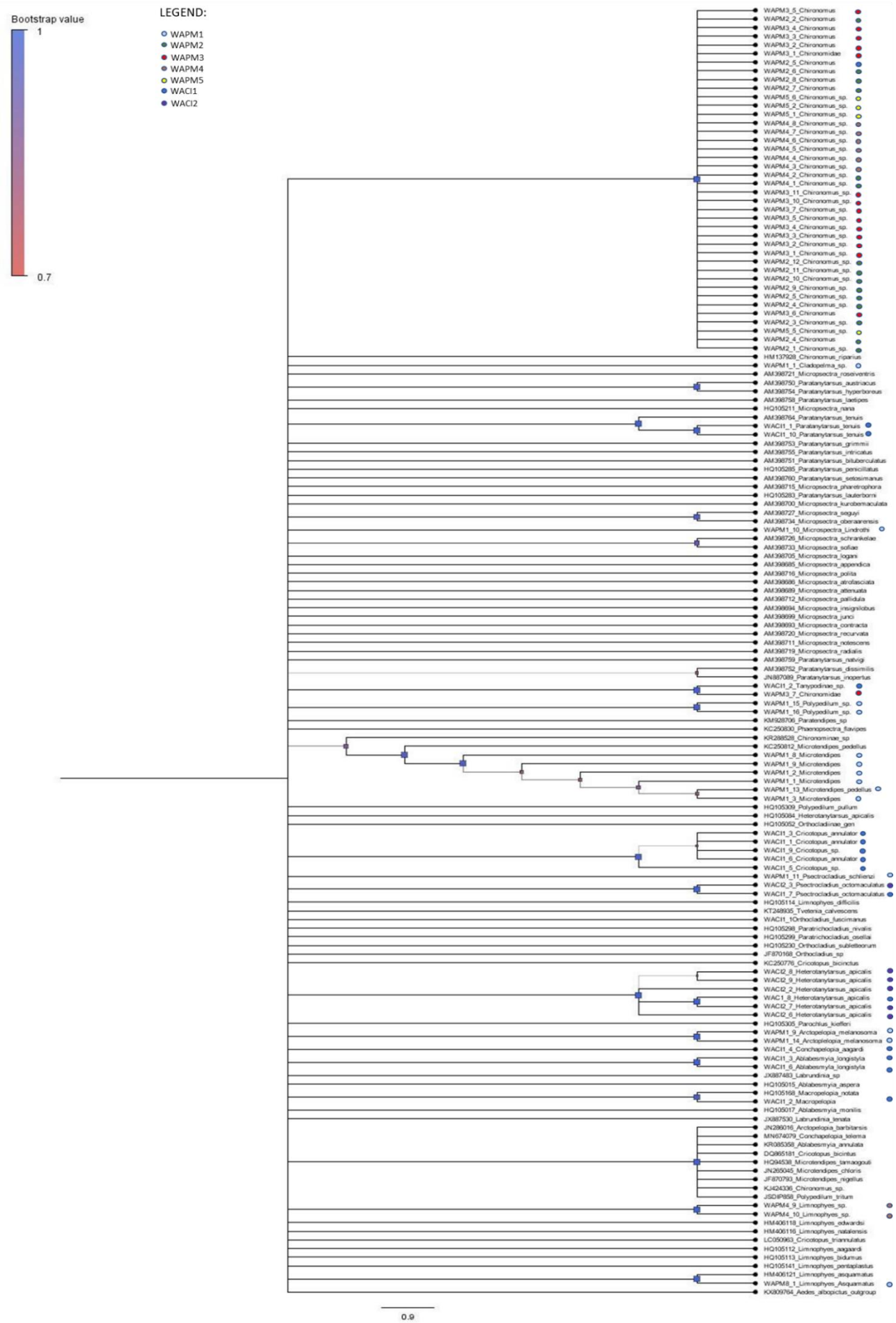
Table 6. Average GenBank identity (using BLAST) of the chironomid larvae from Wales per site and per genus.

	Max Score	Total Score	Query	E-value	Identity (%)
WAPM1	936	912	100	0	97
WAPM2	1000	1043	100	0	99
WAPM3	964	955	100	0	99
WAPM4	1086	1059	100	0	98
WAPM5	1102	1102	100	0	99
WACI1	878	902	100	0	96
WACI2	957	981	100	0	99
Ablabesmyia	1040	1040	92	0	99
Arctopelopia	885	885	96	0	93
Chironomus	923	875	96	0	95
Cladopelma	1048	1048	94	0	99
Cricotopus	863	863	96	0	93
Conchapelopia	780	780	94	0	90
Heterotanytarsus	1000	1000	88	0	100
Limnophyes	957	957	96	0	95
Macropelopia	822	822	100	0	100
Microspectra	1105	1105	96	0	99
Microtendipes	1031	1031	96	0	98
Paratanytarsus	1085	1085	95	0	99
Psectrocladius	1112	1112	96	0	100
Polypedilum	1066	1066	97	0	99
Tanypodinae	778	778	92	0	99
Orthocladius	765	765	100	0	97

PHYLOGENETIC ANALYSIS

All Chironomid sequences from Wales and GenBank were used to build the Neighbour-Joining (NJ) tree (Fig. 7), zoomed sections (see appendix, Fig. S4 A-D). There were two putative genera identified in the suspected “polluted” ponds and 14 in the suspected “non-polluted ponds”. The Neighbour-joining tree was used to observe the genetic diversity within and among ponds. The samples from

WAPM1, and to a lesser extend also from WACI1 and WACI2, were scattered in many branches of the Neighbour-Joining tree, showing that these sites had a high diversity of OTUs. WAPM1 showed 6 different OTUs, while WACI1 and WACI2 showed eight and two numbers of OTUs, respectively. In comparison, sites WAPM2, WAPM3, WAPM4 and WAPM5 showed very little diversity and showed 1, 1, 2 and 1 OTUs, respectively. All the samples from WAPM2, WAPM3, WAPM4 and WAPM5 (apart from WAPM3_7, WAPM4_10 and WAPM4_9) clustered next to a *Chironomus* sequence from GenBank, indicating that the majority of samples from polluted ponds in Parys mountain belonged to this genus, and that some ponds were composed entirely from this genus. Overall, chironomids from Wales were DNA-barcoded reliably (bootstrap-level higher than 70%) to



the genus level.

Figure 7. (continued from previous page). Neighbour-Joining phylogenetic tree of chironomids from Wales and GenBank, rooted with the outgroup Aedes albopictus. Squares on nodes represent bootstrap support values (500 bootstraps), where large squares and blue fill represent high bootstrap support (100 %) and small squares and red fill represent low bootstrap support (0.7 %). Relevant zoomed in sections of the tree can be observed in detail in Appendix Fig. S4(A, B, C, D,).

DNA-BARCODING AND COMMUNITY COMPOSITION

For *Cladopelma* (Fig. 8A), there was just one haplotype in WAPM1. *Ablabesmyia* (Fig. 8B), had just one haplotype in WACI. *Arctopelopia* (Fig. 8C), had just one haplotype in WAPM1. *Cricotopus* (Fig. 8D), had two haplotypes only present in WACI1 distanciated from the GenBank haplotypes. *Heterotanytarsus* (Fig. 8E), had four different haplotypes only present in WACI2. *Conchapelopia* (Fig. 8F), had only 1 haplotype in WACI1. *Microspectra* (Fig. 8G), had just one haplotype in WAPM1. *Microtendipes* (Fig. 8H), had only one haplotype in WAPM1. *Limnophyies* (Fig. 8I) had two haplotypes, one from WAPM3 and one from WAPM4 both distanciated from the GenBank haplotypes. *Polypedilum* (Fig. 8J), had two haplotypes from WAPM1 both distanciated from GenBank samples. *Paratanytarsus* (Fig. 8K), had just one haplotype in Cwm-Idwal. *Orthocladius* (Fig. 8L), had just one haplotype in WACI. *Tanypodinae* (Fig. 8M) had just one haplotype in Cwm-Idwal. *Chironomus* (Fig. 8N) was the most diverse genus with three haplotypes found only in the suspceted polluted ponds: one haplotype was present in in all four the suspected polluted ponds (WAPM2, WAPM3, WAPM4, WAPM5), one hapoltype in WAPM5 and one in WAPM3.

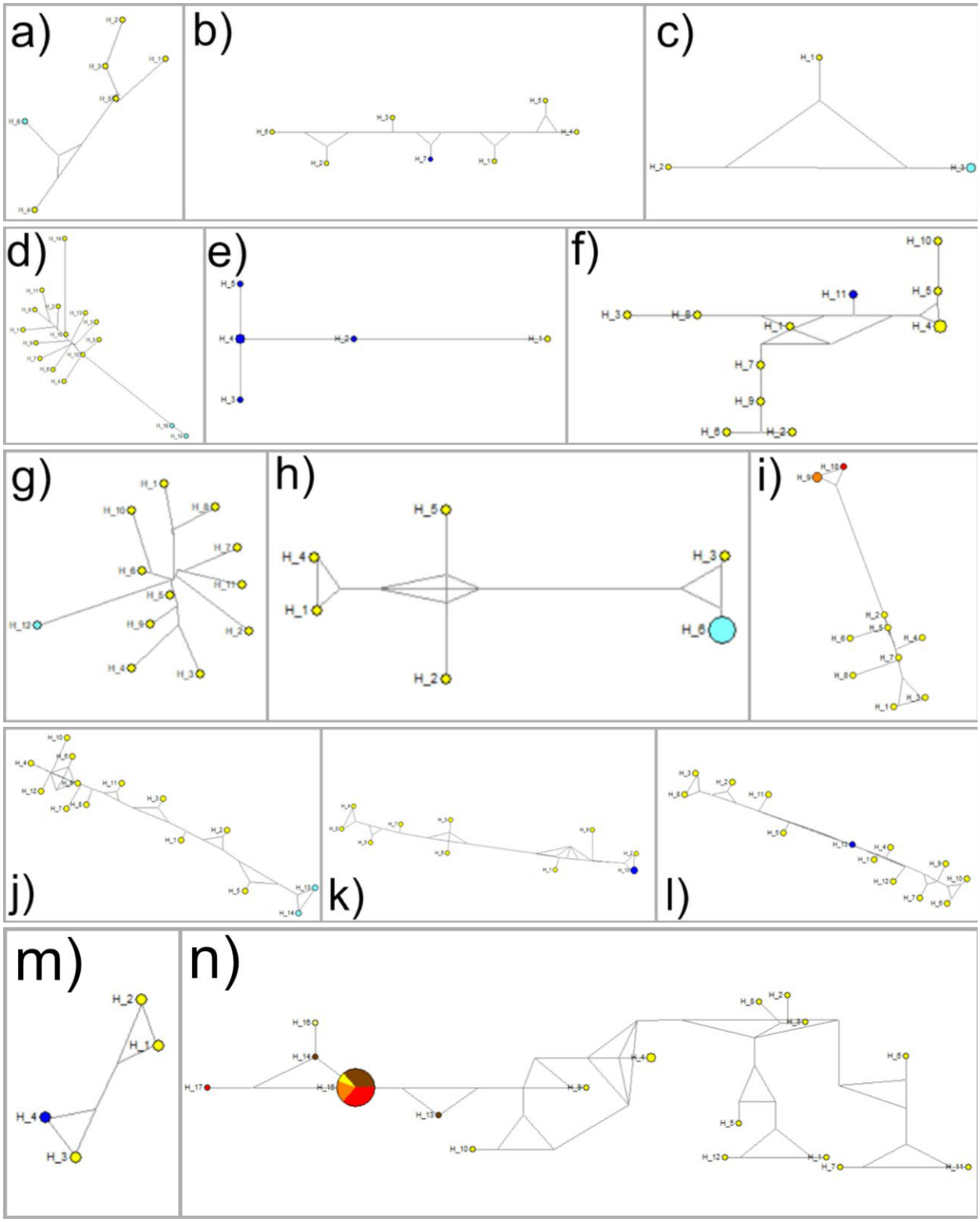


Figure 8. Phylogenetic networks for genera.A) Cladopelma, B) Ablabesmyia, C) Arctopelopia, D) Cricotopus, E) Heterotanytarsus, F) Conchapelopia, G) Microspectra, H) Microtendipes, I) Lymnophyes, J) Polypedilum, K) Paratanytarsus, L) Orthocladius, M) Tanypodinae, N) Chironomus. Colours are respectively: Yellow for GenBank, Light-Blue (WAPM1), Brown (WAPM2), Orange (WAPM3), Red (WAPM4), Light-yellow (WAPM5), Blue (WACI1), Grey (WACI2).

Thirty-one haplotypes plus the outgroup *Aedes albopictus* were identified among Wales population by creating a phylogenic haplotype network (Fig. 9). The haplotypes occurred in distinctive geographical patterns i.e. region that were closer to each other showed more similar haplotypes. Four main clades were found. The first clade was composed by WACI1 and WACI2 from pond WAPM4 (Hap23-26, Hap28-29), the second one formed by WAPM2, WAPM3, WAPM4 and WAPM5 from Parys Mountain (Hap8-11), the third one mainly formed by WAPM1 (Hap 1- 7), WACI1 (Hap12,-3, Hap15-18, Hap20 and Hap30), WAPM4 (Hap13-14 and Hap27) and WACI2l (Hap20). The forth clade was formed by WACI2 (Hap19 and Hap21-22).

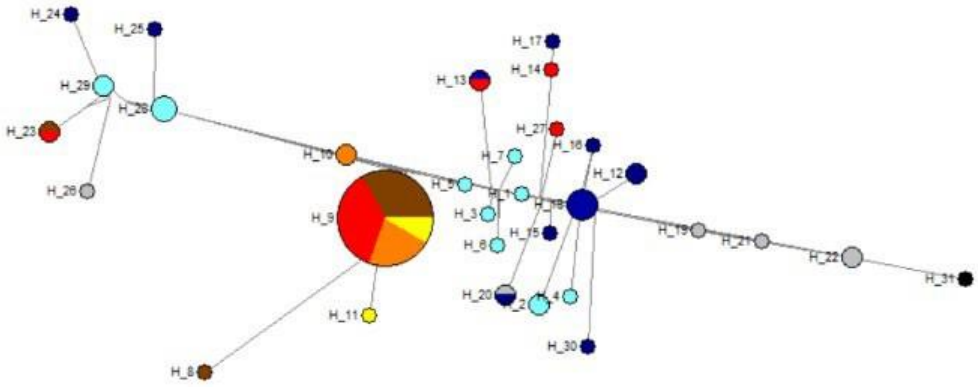


Figure 9. Haplotype network of Chironomidae family based on variation of COI. Circle size represents the relative frequency of a haplotype. Colours are respectively: Light-Blue (WAPM1), Brown (WAPM2), Orange (WAPM3), Red (WAPM4), Yellow (WAPM5), Blue (WACI1), Grey (WACI2).

DNA SEQUENCE POLYMORPHISM

Results from the DNA polymorphism for CO1 between Parys Mountain and Cwm-Idwal ponds (Table 7), show that the number of haplotypes (h) is higher in WAPM1 and WACI1. The haplotype diversity (Hd) ranged from 0 to 1 with a mean of 0.78, while results among the seven specific ponds showed a haplotype diversity mean of 0.53 (Table 8). Results of the average number of nucleotide diversities (K) show low or none (WAPM3) degree of polymorphism within populations in WAPM2, WAPM3, WAPM4 and WAPM5. Measures of nucleotide diversity (π) were low for most populations and zero for ponds WAPM2, WAPM3, WAPM4 and WAPM5 from Parys Mountain. The highest nucleotide diversity values were found in WAPM1 (π=0.15) and WACI1 (π=0.14).

Table 7. Population genetics statistics for global Chironomidae family OTUs (GenBank and Wales). Wales, Parys Mountain, Cwm-Idwal. Number of haplotypes; Hd, haplotype diversity; Pi, nucleotide diversity; S number of segregating sites. WAPM=Parys Mountain; WACI= Cwm-Idwal.

	No. of Sequences	No. of Sites	Nu,ber of Haplotypes (h)	Haplotype diversity (Hd)	Nucleotide Diversity (π)	Averge number of nucleotide differences (K)
Wales and GenBank	615	331	334	0.99	0.44	148.9
GenBank	539	418	298	0.99	0.17	71.9
All Wales	78	351	30	0.78	0.22	80.0
Parys Mountain	57	360	16	0.60	0.16	61.0
Cwm-Idwal	19	417	15	0.97	0.25	105.7
Polluted	54	385	19	0.651	0.18	71.8
Non-polluted	31	363	23	0.97	0.31	114.7

Table 8. Population genetics statistics for global Chironomida family OTUs in all the seven examined ponds present in Wales.

	h	Hd	Pi	S	K
WAPM1	9	0.46	0.15	200	94.17
WAPM2	3	0.46	0.00	17	4.25
WAPM3	2	0.00	0.00	0	0.00
WAPM4	5	0.35	0.00	106	37.68
WAPM5	2	0.50	0.00	3	1.50
WACI1	10	0.96	0.14	176	89.75
WACI2	5	1.00	0.06	98	47.04

F_{ST} values (Table 9) among sites show very low genetic differentiation (< 0.05) between WAPM3-WAPM4 and WAPM3-WAPM5; moderate differentiation (between 0.05 and 0.15) among WAPM2-WAPM5, WAPM4- WAPM5, WAPM2-Wapm4 and Wapm1-WACI1; all the other ponds show high genetic differentiation (> 0.25).

Table 9. Pairwise Fst values of Chironomidae larvae among the seven ponds in Wales (WAPM=Parys Mountain; WACI= Cwm-Idwal).

	WAPM1	WAPM2	WAPM3	WAPM4	WAPM5	WACI1	WACI2
WAPM1		0.59	0.46	0.41	0.58	0.14	0.43
WAPM2			0.37	0.1	0.06	0.61	0.87
WAPM3				0.01	0.03	0.61	0.87
WAPM4					0.09	0.43	0.71
WAPM5						0.6	0.86
WACI1							0.3
WACI2							

The abundance of chironomids varied considerably per site (Fig. 10). The *Chironomus* genus was the most abundant one (n=40) and it was found only in the suspected polluted ponds (WAPM2,

WAPM3, WAPM4, WAPM5). *Limnophyes* (n=3) was also only found in the suspected polluted ponds. All the other genera were found only in the suspected non-polluted ponds (WAPM1, WACI1, WACI2).

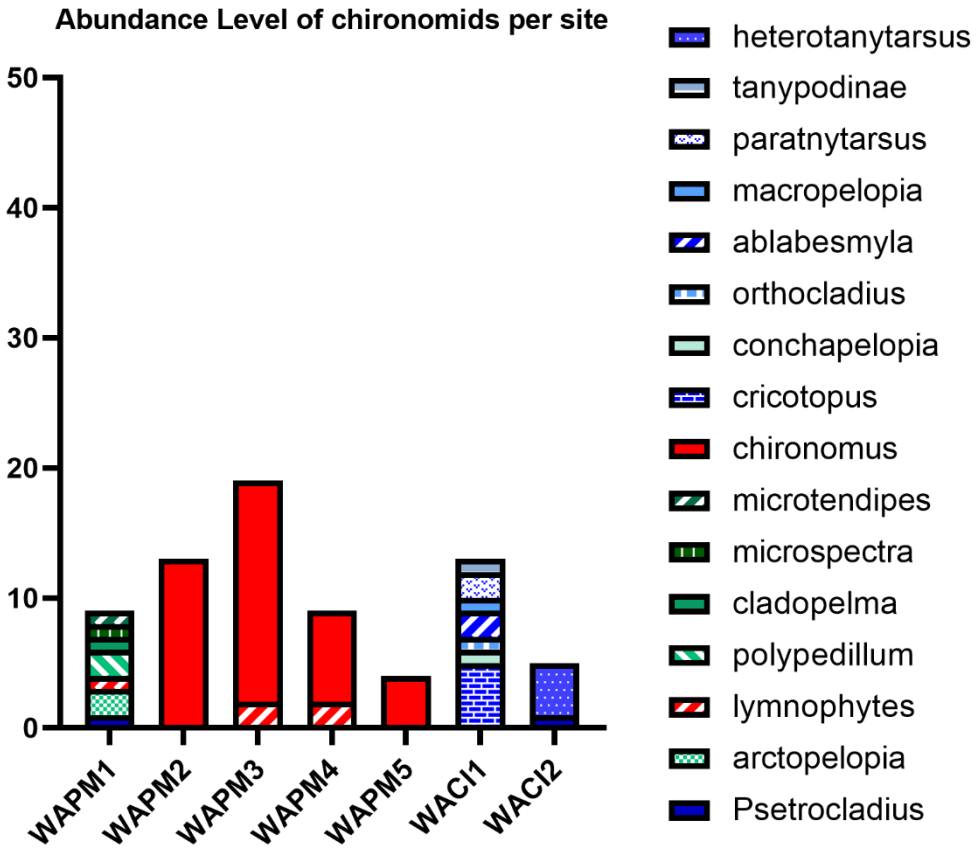


Figure 10. Number (abundance) of chironomid genera for each collection site in Wales.

Simpson's diversity index based on DNA barcoding data for Wales was DS = 0.31. The specific index for each of the seven ponds (Table 10) revealed no diversity in WAPM2 and WAPM5 (DS=0); moderate diversity in WAPM3 and WAPM4 and WACI2 (DS=10-50) and high genetic diversity in WAPM1 and (DS=50-100).

Table 10. Simpson's index across the seven ponds in Wales.

Simpson's index across the ponds in Wales

	WAPM1	WAPM2	WAPM3	WAPM4	WAPM5	WACI1	WACI2
No. of Individuals	13	14	15	10	4	15	7
Simpson's Index	0,72	0	0,12	0,32	0	0,82	0,24

The Bray-Curtis similarity analysis showed two different clusters (Fig. 11), one cluster formed by the suspected polluted ponds (WAPM2, WAPM3, WAPM4 and WAPM5) and one cluster formed by the suspected non-polluted ponds (WAPM1, WACI1, and WACI2) with an average of 57% of similarity. However, pairwise Bray-Curtis similarity was not very high, with about 10% between WACI1 and WACI2, 5% between WAPM1 and the two ponds from Cwm-Idwal. However, a highest similarity was found between the suspected polluted ponds (>80% similarity), and with WAPM2 and WAPM5 being identical based on chironomid larvae composition.

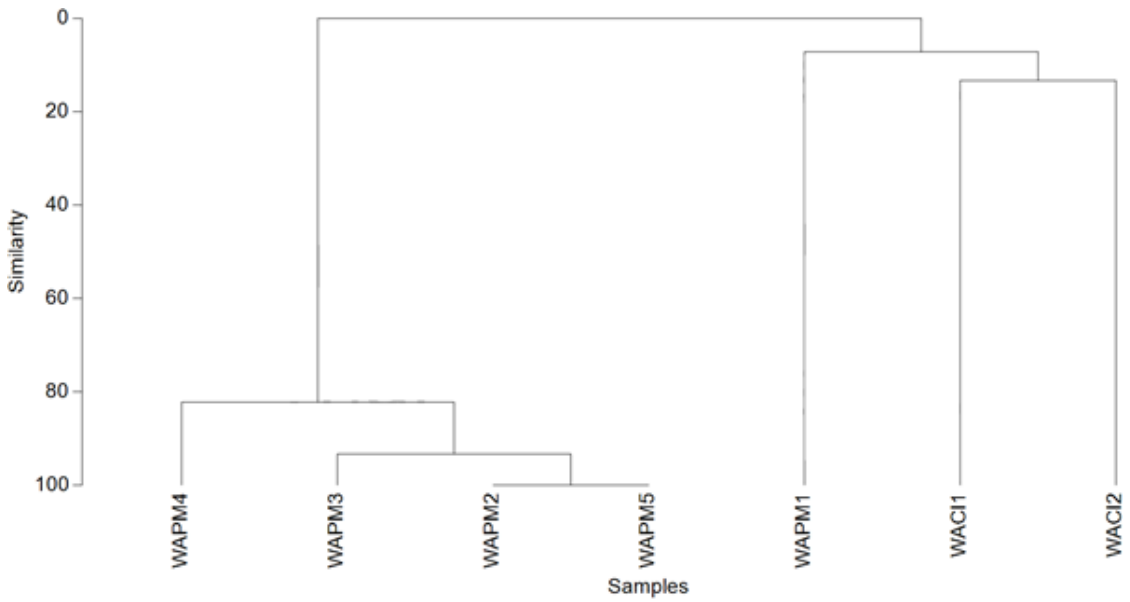


Figure 11. Bray-Curtis similarity among collection sites.

Results from the 2D MSD (Fig. 12), showed three main different groupings: one formed of WACI1 and WACI2, formed WAPM1 and one formed by WAPM2, WAPM3, WAPM4 and WAPM5. Even if WAPM1 resulted in a separate group, it resulted more closely related to WACI1 and WACI2 than the other Parys Mountain ponds.

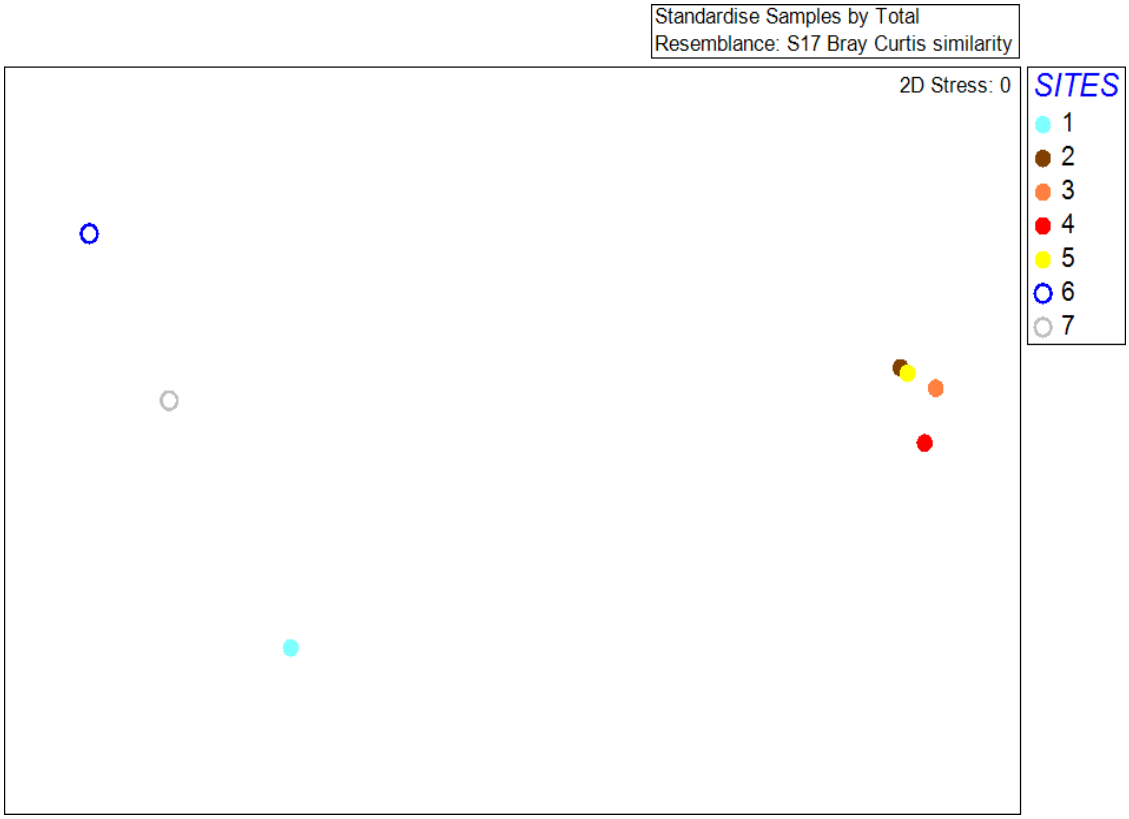


Figure 12. 2D MDS plot (non-metric) of the seven ponds analysed in Wales. Stress value: 0; zero-adjusted Bray-Curtis similarity.

The global ANOSIM test (Table 10) showed there were significant differences in community structure among the seven ponds (R=0.35, p=0.001).

Table 10 R and P values obtained with ANOSIM test. Significant differences ($p<0.005$**; $R>0.30$) were highlighted in bold.

R value							
	WAPM1	WAPM2	WAPM3	WAPM4	WAPM5	WACI1	WACI2
WAPM1		**0.63**	**0.61**	**0.37**	0.27	0.16	**0.31**
WAPM2			0	0.08	0	**0.51**	**0.94**
WAPM3				0.05	0	**0.5**	**0.85**
WAPM4					0	0.27	**0.66**
WAPM5						0.18	**0.77**
WACI1							0.16
WACI2							

P value							
	WAPM1	WAPM2	WAPM3	WAPM4	WAPM5	WACI1	WACI2
WAPM1		**0.001**	**0.001**	**0.001**	**0.003**	**0.001**	**0.001**
WAPM2			1	0.014	1	**0.001**	**0.001**
WAPM3				0.32	1	**0.001**	**0.001**
WAPM4					1	**0.001**	**0.001**
WAPM5						**0.001**	0.3
WACI1							0.17
WACI2							

Discussion

DNA-barcoding

The study was designed to use DNA-barcoding to identify chironomid larvae in polluted and non-polluted ponds in Wales and assess whether there were differences in community structure and diversity based on environmental characteristics of sediments and water. Within Parys Mountain, one pond was considered to be non-polluted and four others were considered polluted because they were used as desiccation ponds for extraction of metals. The two ponds from Cwm-Idwal were considered to be non-polluted because they are fed by mountain streams in an area with little human activities.

DNA sequence variation in these population-level surveys is often quantified using data from haploid loci, with the mitochondrial gene cytochrome c oxidase subunit 1 (CO1) being the most

frequently sequenced in animals. The popularity of CO1 stems from the availability of several sets of conserved PCR primers (for example, Folmer et al. 1994) and the dual purpose of cox1 as a molecular marker for both intra-specific variation and species identification (Bucklin et al. 2011). Overall, the DNA-barcoding results were consistent with previous studies (Sweeney et al. 2011; Hebert et al. 2005; Ekrem et al. 2007; Carew et al. 2007) showing that DNA-barcoding is a powerful tool in providing quick genera and species identification by using a short standardised gene region as an internal species tag. Moreover, it was demonstrated that mtDNA CO1 genes was useful in chironomid larvae identification. In this research, 78 out of 120 samples were successfully amplified and barcoded to GenBank data. The average identity (98.14%) obtained in the seven ponds in Wales was below 100%. However, being the percent identity close to 100% on each site and the bootstrap levels of the phylogenetic tree, the putative genus-level identification was considered adequate. Morphological identification would be required in order to confirm the specific identity of the samples (Shen et al. 2016). Results of subfamily groups were consistent with previous reports (Ekrem et al. 2007; Cranston et al. 2012; Sari et al. 2015). The phylogenetic tree was used as a complement to the BLAST research showing an evolutionary relationship among the samples from Wales and GenBank. The length of the branches represent genetic distance (Yuan et al., 2014). As the chironomids belong to the same family, no significant genetic difference was expected between species. This was confirmed by the similarity in branch lengths. Results from the phylogenetic tree showed that all the Chironomus specie from the suspected polluted ponds clustered together. However, the Chironomus sequence retrieved from GenBank didn't cluster with them. The sequence from the suspected polluted pond WAPM3_7 identified as chironomidae was found clustered with the tanypodinae species from the suspected non-polluted pond WACI1, therefore there is a possibility that even though the DNA-barcoding wasn't able to identify the species but only the family, the sample WAPM3_7 could be identified as a tanypodinae. Also, the lymnophytes spieces from the suspected pollute ponds was found clustered distantly from the other species from the polluted ponds. This could suggest that even though the suspected polluted ponds showed a low internal genetic difference compared to the

suspected non-polluted ones, there is still the possibility to find a small variation. All the genera from the suspected non-polluted ponds are found in several branches throughout the phylogenetic tree. This again, support the fact that there is a higher variation amongst the suspected non-polluted ponds compared to the polluted ones.

The bootstrap support in the phylogenetic tree external nodes was mostly high showing high confidence levels in the relationships between samples from Wales and GenBank. However, the internal node showed low bootstrap level this could be probably due to difficulties in the alignment (Marescaux et al. 2013).

COMMUNITY COMPOSITION

The haplotype diversity analysis showed a generally high diversity in the seven ponds in Wales (Hd=0.78). According to Hobbs et al. (2013), the haplotype diversity within invertebrate species is considered low when smaller than 0.5. The higher haplotype diversity values were found in the two suspected non-polluted ponds from Cwm-Idwal: WACI1(0.96) and WACI2 (1.00). However, the haplotype diversity is affected by sample size (Goodall-Copestake et al. 2012). The nucleotide diversity was taken into consideration when looking at the internal genetic diversity within the ponds. Readily assess the wider significance of population-level CO1,h and p estimates are often hard to find as there is no published summary of this data. Specific comparisons between h or p from different studies are rare and context-dependent, involving species sharing taxonomic/ecological characteristics (for example, Kim et al. 2009) or population-level samples with similar levels of diversity (for example, Goodall-Copestake et al. 2010). Furthermore, similar values of h or p may be referred to as low in one publication but high in another. Measures of nucleotide diversity (π) were low for most populations and zero for ponds 2,3,4 and 5 from Parys Mountain. The highest nucleotide diversity values were found in WAPM1 (π =0.15) and WACI1 (π =0.14) indicating that between the seven ponds, the suspected non-polluted ones have the highest genetic diversity.

DNA polymorphic sites were analysed by looking at the parsimony sites within 2, 3 and 4 variants and finally the gene flow and genetic differentiation were analysed by looking at the Wright's fixation index (F_{ST}) (Wright 1951). F_{ST} values, representing genetic differentiation, can be low (> 0.05), moderate (between 0.05 and 0.15) or high (>0.25) (Balloux et al. 2002). F_{ST} among sites showed low genetic differentiation between WAPM3-WAPM3 and WAPM3-WAPM5 (< 0.05); moderate differentiation among WAPM2-WAPM4, WAPM2-WAPM5, WAPM4-WAPM5 and WAPM1-WACI1 (between 0.05 and 0.15); all the other ponds show high genetic differentiation (> 0.25). For the interpretation of F_{ST}, it has been suggested that a value lying in the range 0–0.05 indicates little genetic differentiation; a value between 0.05 and 0.15, moderate differentiation; a value between 0.15 and 0.25, great differentiation; and values above 0.25, very great genetic differentiation (Wright 1978; Hartl and Clark 1997). Indeed, an F_{ST} of 0.05 will generally be considered as reasonably low and investigators may interpret that structuring between sub-populations is weak. While such an interpretation may turn out to be correct, it may also not be representative at all of the real population differentiation (Balloux et al. 2002).

The non-biting midge family Chironomidae is divided into 11 subfamilies: Aphroteniinae, Buchonomyiinae, Chilenomyiinae, Chironominae, Diamesinae, Orthocladiinae, Podonominae, Tanypodinae, Telmatogetoninae and Usambaromyiinae (Leung et al. 2011). From these, only three subfamilies have been identified in Wales: Chironominae, Orthocladiinae and Tnaypodinae. The genera *Chironomus, Cladopelma, Microspectra, Microtendipes, Paratanytarsus* and *Polypedilum* belong to the subfamily Chironominae. The genera *Cricotopus, Heterotanytarsus, Limnophyes, Psectrocladius* and *Orthocladius* belong to the subfamily Orthocladiinae. The genera *Ablabesmyia, Arctopelopia,Conchapelopia, Tanypodinae and Macropelopia* belong to the subfamily Tanypodinae.

Based on taxonomical identification, the suspected non-polluted ponds were found very different in composition. The higher Bray-Curtis similarity between the suspected non-polluted ponds was

expected. Also, the higher similarity between the two suspected non-polluted ponds from Cwm-Idwal and the suspected non-polluted ponds from Parys mountain was expected even if geographically distant. WACI1 had the highest richness and abundance of chironomids also resulting in the higher Simpson's diversity index. The first of the heterogeneity indices used in ecology was proposed by Simpson. This index measures the probability that two individuals selected at random from a sample will belong to the same species (Peet 1974). It would be expected that further sampling based on different seasons would show different richness and diversity as one of the most important factors controlling the distributions of chironomids in lotic environments are temperature and flow regime, which also control the availability of food and substratum to a large extent (Lindegaard, 1995).

The 2D MSD plot re-confirmed what the Bray-Curtis similarity was showing as the suspected polluted ponds grouped together where WAPM1 and showed a higher similarity and the suspected non-polluted ponds showed as a distant cluster. The 2D-MSD plot showed a 2D stress level of 0. According to Hout et al. (2013), *stress function* is calculated (the particular function varies across different scaling algorithms), which measures the agreement between the estimated distances and the input proximities. Lower stress values indicate a better fit; thus, the algorithms attempt to increase the fidelity to the input data by minimizing this stress function (Hout et al. 2013). The global ANOSIM test showed there were significant differences in community structure among the seven ponds (R=0.35, p=0.001). An R value close to "1.0" suggests dissimilarity between groups while an R value close to "0" suggests an even distribution of high and low ranks within and between groups. A R=0.35 show a difference in the composition of the ponds although not extremely separated.

ENVIRONMENTAL CONDITIONS

The environmental and magnetic susceptibility analysis was done on water and sediments samples to determine the environmental conditions and potential presence of heavy metals (pollution) in the ponds.

The Magnetic susceptibility and the IRM were calculated to calculate the amount of magnetic metal in response to a magnetic field (Murdock et al. 2013). Overall, there was no particular sign of strong magnetic susceptibility values compared to the others. All the environmental conditions were statistically analysed. The analysis show a significant difference in the pH, salinity and dissolved solids between polluted and non-polluted ponds. However, it was expected that the pH, dissolved oxygen and salinity would result significantly different as according to literature (McKie et al. 2005; Nyman et al. 2005; Panatta et al. 2007; Kranzfelder et al. 2017), they can have an influence on the Chironomid larvae.

Chironomid larvae can tolerate high levels of salinity; and they usually represent more than 70% of the macroinvertebrate population in hypersaline water (Shadrin et al. 2017). Salinity data analysed showed higher salinity levels in WAPM2,3,4 and 5 than in WAPM1 and WACI1 and 2 leading to assume that only the genera *Chironomus* and *Lymnophites* can tolerate these high salinity values. The fact that the suspected polluted ponds had higher salinity levels, could be due to the high salinity levels are present in freshwater systems when mineralisation of water is determined by the degradation of inorganic materials (such as heavy metals) in the water (Zinchenko et al. 2012). Townsed et al. (1983) offered a convincing argument that acidity is the main factor in stream community composition. Low pH can be caused by the presence of heavy metals in the water (Henrikson et al. 1985). According to Bell et al. (1970) and Orendt (1999), most of the chironomid species are not able to complete their life cycle under a certain level of pH (5.0). As data collected in the field showed, pH levels in WAPM2, WAPM3, WAPM4 and WAPM5 were low. As shown in Henriskon et al. (1985), there seem to be a correlation between low pH and low oxygen levels in the water. On one hand, dissolved oxygen levels collected on the field, show no correlation with pH. Moreover, oxygen levels measured in the field were approximately similar among all the ponds but slightly higher in the suspected non-polluted ponds. On the other hand, even though dissolved oxygen levels were not significantly low in WAPM2, WAPM3, WAPM4 and WAPM5, several studies have shown that *Chironomus* species are the best adapted to low oxygen levels (Lagauzere et al. 2009; Lee

et al. 2006; Lenat 1993). It has been suggested that *Chironomus* is best adapted to low oxygen level in the water for two reasons: according to Hamburger et al. (1995), *Chironomus* larvae are able to survive in low oxygen concentration levels, or even anoxia, because they can produce large glycogen stores which may be converted to ethanol or lactic acid during anaerobic metabolism; another reason is that they possess haemoglobin during their larvae stage, which gives them the particular red colouration and help them cope with low oxygen levels (Ha et al. 2008).

It seems that because of the history of the site (Parys Mountain is an old site used for mining), heavy metals caused differentiation in environmental conditions in the water, affecting the chironomid larvae. The lowered pH-value per se might, of course, be of importance. A lowered pH value could also be a releasing factor for other important mechanisms, biotic as well as abiotic. Examples of abiotic factors include increases in heavy metals which have been reported from acidified lakes (Aimer et al. 1978). Data obtained from the concentration of the heavy metals from the sediments in the ponds collected in Parys Mountain, resulted consistent with previous researches (Wilson and Pyatt, 2007; Rios et al. 2007; Marsay 2018). By comparing the concentration levels of heavy metals obtained from the pond sediments (Table 3), to the maximum concentration levels allowed near agricultural activities from DEFRA, respectively: 565 mg/kg for Copper; 270 mg/kg for Iron; 221 mg/kg for Lead and 802 mg/kg for Zinc, is possible to notice that for all the suspected polluted ponds in Parys mountain, the heavy metals' concentration levels exceed the maximum allowed ones (Vincent and Passant, 2013).

FUTURE DIRECTIONS AND CONSIDERATIONS

Based on the mtDNA diversity found across the seven ponds in Wales, it can be assumed that there is a difference in the chironomid genetic and community composition caused by some environmental factors which is worth exploring further. Further research could use a broader number of samples and collection could be performed at different periods of time in order to fully understand how the

biodiversity and the generic compositions of the chironomids changes. There are still broad knowledge gas in chironomid species in the United Kingdom and therefore more studies are needed to fill the gaps. With the rapid development and lowered costs of modern molecular techniques, integrative taxonomy combining morphological and molecular data should be more available to taxonomists. This will certainly provide the scientific community more accurate measures of the diversity in one of the most abundant and species rich groups of aquatic insects.

A major benefit of, and rationale for, DNA barcoding lies in its cost-effectiveness for species identification, especially in ambitious bioinventory and biomonitoring programs (Smith et al. 2006). Cost is only one criterion in evaluating the utility of a taxonomic support system for biodiversity research. Speed, reliability, and accessibility are just as important, and it is believed that DNA barcoding excels in these areas. Some of the limitations of DNA barcoding are that the DNA barcodes are only assigned a putative name, which still needs to be verified by experts, it depends on having a good 'library' available for the taxonomic group of interest and for the geographical region being sampled, and it currently relies on the understanding of mitochondrial inheritance of a single molecular marker (Frezal and Lebois 2008).

Conclusions

This study had four main aims: 1) to characterise the levels of heavy metals pollution in ponds and their possible influence on the Chironomids, 2) to identify non-biting midge (Chironomids) larvae in ponds in Wales through DNA barcoding, 3) to characterise the genetic diversity and genetic structure of non-biting midges among ponds in Wales, and 4) to characterise the community structure of non-biting midges in Wales. The DNA sequencing of chironomid larvae in this research was useful for providing a first picture of the genetic diversity and structure in Wales as there are not many DNA barcoding studies which report richness, diversity and community structure from the DNA barcodes.

The DNA-based approach used in this project allowed to assign with high level of confidence

each mtDNA sequence a putative name without the use of morphological identification which often is time-consuming. The identity results were based on the best match in GenBank and then confirmed by the phylogenetic methods. Even though there is no certainty that the heavy metals pollution in ponds have an influence on the chironomids, there seem to be a correlation between the environmental characteristics of ponds and the chironomid larvae diversity and community structure as well as with the chironomid larvae genetic diversity and structure.

Significant heavy metal concentrations were detected at the mine workings at Parys Mountain. The presence of heavy metals in the sediments may have caused the acidic mine drainage (AMD) affecting the environment in different ways such as lowering the pH. This lower pH makes the environment uninhabitable for many flora and fauna but those that can survive the lower pH are threatened by toxic metals, which are more bio-available in low pH environments. In conclusion, this project was able to identify the non-biting midges larvae through DNA-barcoding although morphological identification would add an extra amount of certainty. The levels of heavy metals pollution in ponds and their influence on the Chironomids larvae was established confirming that the suspected polluted ponds were indeed over the threshold of heavy metals' concentration permitted from DEFRA. Finally, we were able to characterise the genetic diversity and genetic structure of non-biting midges among ponds in Wales, and to characterise their community structure.

Acknowledgements

I would like to thank my supervisors Dr Rodrigo Vega and Dr Gregg McIntosh for their patience, for always encouraging my research and for allowing me to grow as a research scientist. Anglesey Mining plc allowed me to work in their site. I also acknowledge Dr Naomi Rintoul for her guidance and support throughout this research project, and Joe Burman, John Hills, Owain Williams, Naomi Beddoe, Hannah Shiling and Jaimie Morris for comments, support and discussion throughout this project, and other members of the laboratory staff at Canterbury Christ Church University for their help in the laboratory during this time.

References

Abdi, H., Williams, L. (2010) Principal Component Analysis. John Wiley &Sons inc.

Almeida C., Quintar S., González P. and Mallea M., (2008) "Assessment of irrigation water quality". A proposal of a quality profile. Environ. Monit. Assess. Vol.142, No. 1–3, pp149–152.

Almer, B., Dickson, W., Ekstr?m, C. and H?rnstr?m, E.1978. Sulfur pollution and the aquatic ecosystem.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Anglesey Mining plc. (2010) *Parys mountain a special place.* [online].

Arillo, A., Buffagni, A., Camusso, M., Ciannarella, R., Crosa, G., Falugi, C., Galassi, S., Guzzella, L., Lopez, A., Mingazzini, M., Pagnotta, R., Patrolecco, L., Tartari, G. and Valsecchi, S. (2003). Quality assessment of bed sediments of the Po River (Italy). *Water Research*, [online] 37(3), pp.501–518.

Armitage, P., Cranston, P., Pinder. S. (1995) "The Chironomidae: Biology and Ecology of Non-Biting Midges". Chapman & Hall, London.

Balloux, F., Lugon-Moulin, N. (2002) "The estimation of population differentiation with microsatellite markers", *Molecular ecology,* 11, pp. 155-165.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution, 16, 37–48.

Bartam, J. (1996). Water Quality Monitoring. A practical guide to the design and implementation of freshwater quality studies and monitoring programmes.

Battarbee R, Anderson N, Appleby P, Flower RJ, Fritz S, Haworth E, Higgit S, Jones V, Kreiser A, Munro MA, Natkanski J, Oldfield F,Patrick ST, Richardson N, Rippey B, Stevenson AC (1988). Lake Acidification in The United Kingdom, ENSIS, London.

Bayat, A. (2002) Bioinformatics. Clinical review

Beaty, B., Black, W., Carlson, J., Clements, W., DuTeau, N., Harrahy, E., Nuckols, J., Olson, K., Rayms-Keller, A. (1998) "Molecular and Genetic Ecotoxicologic Approaches to Aquatic Environmental Bioreporting", *Environmental Health Perspectives*, 106(6), pp. 1395-1407.

Bell, H. (1970) "Effects of pH on the life cycle of the midge *Tanytarsus dissimilis*", *Entomological Society of Canada*, 102, PP. 636-639.

Bickham, J. (2019). *Effects of chemical contaminants on genetic diversity in natural populations: implications for biomonitoring and ecotoxicology*.

Bucklin A, Steinke D, Blanco-Bercial L (2011). DNA Barcoding of Marine Metazoa. Annu Rev Mar Sci 3: 471–508.

Burgess KS, Fazekas AJ, Kesanakurti PR et al. (2011) Discrimi- nating plant species in a local temperate flora using the rbcL+matK DNA barcode. Methods in Ecology and Evolution, 2, 333–340.

Carew, M., Pettigrove, V., Cox, R., Hoffman, A. (2007) "DNA identification of urban Tanytarsini chironomids (Diptera: Chironomidae)", *Journal of the North American Benthological Society*, 26(4), pp. 587-600.

Chao, T.T., 1984. Use of partial extraction techniques in geochemical exploration. J. Geochem. Explor. 20, 101–135.

Chernicoff, S. (1984) "Using Isotropic Magnetic Susceptibility To Delineate Glacial Tills", *The journal of Geology*, 92(1), pp. 113-118.

Clarke, R., Somerfield, P., Chapman, M. (2006) "On resemblance measures for ecological studies, including taxonomic dissimilarities and zero-adjusted Bray-Curtis coefficient for denuded assemblages", *Journal of Experimental Marine Biology and Ecology*, 330, pp. 55-80.

Clement M, Posada D, Crandall KA (2000) TCS: a computer pro- gram to estimate gene genealogies. Molecular ecology, 9, 1657– 1659.

Cranston, P. (2006) "A new genus and species of Chironominae (Diptera: Chironomidae) with wood-mining larvae", *Australian Journal of Entomology*, 45, pp. 227-234.

Dalziel, A.C., Rogers, S.M. and schulte, P.M. (2009). Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Molecular Ecology*, [online] 18(24), pp.4997–5017.

Darwin, C., 1842. Notes on the effects produced by the ancient glaciers of Caernavonshire, and on boulders transported by floating ice. Philos. Mag. 3, 180–188.

De Haas, E., Wagner, C., Koelmans, A., Kraak, M., Admiraal, W. (2006) "Habitat selection by chironomid larvae: fast growth requires fast food", *Journal of Animal Ecology*, 75, pp. 148-155.

Debels P., Figueroa R., Urrutia R., Barra R. and Niell X., (2005) "Evaluation of water quality in the Chilla´ n River (Central Chile) using physicochemical parameters and a modified water quality index". Environ. Monit. Assess. Vol.110, pp301–322.

Dillon, P. (1985) "Chironomid Larval Size and Case Presence Influence Capture Success Achieved by Dragonfly Larvae", *Freshwater Invertebrate Biology*, 4(1), pp.22-29.

Ekrem, T., Stur, E., Hebert, P. (2010) "Females do count: Documenting Chironomidae (Diptera) species diversity using DNA barcoding", *Organisms Diversity & Evolution*, 10, pp. 397-408.

Environmental Mining Council of BC. (2000). Acid Mine Drainage: Mining and Water Pollution Issues in BC. http://www.miningwatch.ca/updir/amd.pdf

Escritt, E.A., 1971. Plumbing the depths of Idwal's moraines. Geogr. Mag. 44, 52–55. Fisher, P.F.,Tate, N.J., 2006. Causes and consequences of error in digital elevation models. Prog. Phys. Geogr. 30, 467–489.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics, 1, 47–50. Fagundes.

Farkas, A., Erratico, C. and Viganò, L. (2007). Assessment of the environmental significance of heavy metal pollution in surficial sediments of the River Po. *Chemosphere*, [online] 68(4), pp.761– 768.

Fitzsimmons, J. (2015) Analytical method validation: ICP-OES. The royal society of chemistry.

Floyd R, Eyualem A, Papert A, Blaxter ML (2002) Molecular barcodes for soil nematode identification. Molecular Ecology, 11, 839–850.

Folmer, O., Black, M., Hoeah, W., Lutz, R., Vrijenhoek, R. (1994) "DNA primers for amplification of mithocondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates", *Molecular Marine Biology*, 3(5), pp. 294-299.

Foote, B. (1987) "Chironomidae (Chironomoidea)", Immature *Insects*, (2), pp. 762-764.

Friedheim, S. (2015). *Comparison of SpeciesIdentification Methods: DNA Barcoding versus Morphological Taxonomy*.

Galan, M., Pagès, M. and Cosson, J.-F. (2012). Next-Generation Sequencing for Rodent Barcoding: Species Identification from Fresh, Degraded and Environmental Samples. *PLoS ONE*, [online] 7(11).

Galimberti, A., Romano, D.F., Genchi, M., Paoloni, D., Vercillo, F., Bizzarri, L., Sassera, D., Bandi, C., Genchi, C., Ragni, B. and Casiraghi, M. (2012). Integrative taxonomy at work: DNA barcoding of taeniids harboured by wild and domestic cats. *Mol. Ecol. Resour.* 12 (3), pp. 403–413.

Garbarino JR, Hayes H, Roth D, Antweider R, Brinton TI, Taylor H (1995). Contaminants in the Mississippi River, U. S. Geological Survey Circular 1133, Virginia, U.S.A. (www.pubs.usgs.gov/cir-c/circ1133/).

GESAMP 1988 Report of the Eighteenth Session, Paris 11-15 April 1988. GESAMP Reports and Studies No. 33, United Nations Educational, Scientific and Cultural Organization, Paris.

Gilbert, R. (2019). *Statistical Methods for Environmental Pollution Monitoring*. [online] Osti.gov.

Goodall-Copestake, W., Tarling, G. and Murphy, E. (2012) “On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene cox1 in animals”, *Heredity*, 109(1), pp.50-56.

Ha, M., Choi, J. (2008) “Effects of environmental contaminants on haemoglobin of larvae of aquatic midge, *Chironomus riparius* (Diptera: Chironomidae): A potential biomarker for ecotoxicity monitoring”, *Chemosphere*, 71, pp. 1928-1936.

Hall, T. (1998) “BioEdit: a user-friendly biological sequence alignment editor and analysis pro- gram for Windows 95/98/NT. Nucleic Acids Symposium Series”, *Advances in Microbiology,* 41, pp. 95–98.

Hamburger, K., Dall, P., Lindegaard, C. (1995) “Effects of oxygen deficiency on survival and glycogen content of *Chironomus anthracinus* (Diptera, Chironomidae) under laboratory and field conditions”, *Hydrobiologia*, 297, pp. 187-200.

Hebert, P., Gregory, T. and Savolainen, V. (2005) “The Promise of DNA Barcoding for Taxonomy”, *Systematic Biology*, 54(5), pp.852-859.

Henrikson, L., Oscarsons, H. (1985) “History of the Acidified Lake Gardsjon: The Development of Chironomids”, *Ecological Bulletins,* 37, pp. 58-63.

Hill, M. (2010). Understanding environmental pollution. cambridge University press.

Hobbs, J., van Herwerden, L., Jerry, D., Jones, G. and Munday, P., 2013. High Genetic Diversity in Geographically Remote Populations of Endemic and Widespread Coral Reef Angelfishes (genus: Centropyge). *Diversity*, 5(1), pp.39-50.

Hout, M., Papesh, M., Goldinger, S. (2013). “Multidimensional Scaling”, *NIH Public access*, 4(1), pp.93-103.

Hutton M, Symon C (1986). The Quantities of Cadmium, Lead, Mercury and Arsenic Entering the U.K. Environment from Human Activities. Sci. Total Environ. 57: 129-150.

Johnson, D., Ambrose, S., Bassett, T., Bowen, M., Crummey, D., Isaacson, J., Johnsons, D., Lamb,

P., Saul, M and Winter-Nelson, A. (1997). “Meanings of Environmental Terms*”, Journal of Environental Quality,* 25 (3).

J. O. Duruibe and M. . O. C. Ogwuegbu, “Heavy metal pollution and human biotoxic effects,” International Journal of Physical Sciences, vol. 2(5), pp. 112-118, 2007. [online] Available at: http://www.sciepub.com/reference/145718 [Accessed 26 Aug. 2019].

Karataglis, S. (1982) “ Comnbined Tolerance TO copper, Zine and Lead by Populations of Agrostis Tenuis”, *Nordic Society Oikos,* 38(2), pp. 234-241.

Kazi, T., Arain, M., Jamali, M., Jalbani, N., Afridi, H., Sarfraz, R., Baig, J. and Shah, A. (2019). *Assessment of water quality of polluted lake using multivariate statistical techniques: A case study*.

Kim, S., Song, K., Ree, H., Kim, W. (2012) “A DNA Barcvode Library for Korean Chironomidae (Insecta: Diptera) and indexes for Defining Barcode Gap”, *Molecules and cells,* 33, pp. 9-17.

Kirk, H. and Freeland, J.R. (2011). Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. *International Journal of Molecular Sciences*, [online] 12(6), pp.3966–3988.

Kirsten,M; Collar, J., Dolcemascolo, P. (2010) "Using Molecular Techniques to answer ecological questions. *Nature Education*.

Koklu, R., Sengorur, B. and Topal, B. (2009). Water Quality Assessment Using Multivariate Statistical Methods—A Case Study: Melen River System (Turkey). *Water Resources Management*, [online] 24(5), pp.959–978.

Kranzfelder, P., Ekrem, T., Stur, E. (2017) “DNA Barcoding for Species Identification of Insect Skins: A Test on Chironomidae (Diptera) Pupal Exuviae”, *Journal of insect Science,* 17(6), pp. 1-7.

Lagauzere, S., Pischedda, L., Cuny, P., Gilbert, F., Stora, G., Bomzom, J. (2009) “Influence of *Chironomus riparius* (Diptera, Chironomidae) and *Tubex tubifex* (Annelida, Oligochaeta) on oxygen uptake by sediments. Consequences of uranium contamination”, *Environmental Pollution*, 157, pp. 1234-1242.

Lambert, D. (1995) The new science of Molecular Ecology. New Zealand Journal of Ecology.

Lee, S., Lee, S., Park, C., Choi, J. (2006) " Expression of heat shock protein and hemoglobin genes in *Chironomus tentans* (Dipetera, chironomidae) larvae exposed to various environmental pollutants: A potential biomarker of freshwater monitoring", *Chemosphere,* 65, pp. 1074-1081.

Lenat, D. (1993) "Using Mentum Deformities of Chironomus Larvae to Evaluate the Effects of Toxicity and Organic Loading in Streams", *Journal of the North American Benthological Society,* 12(3), pp. 265-269.

Lenntech Water Treatment and Air Purification (2004). Water Treat- ment, Published by Lenntech, Rotterdamseweg, Netherlands (www.excelwater.com/thp/filters/Water-Purification.htm).
Levin, S. (2010) The evolution of ecology. The chronicle of higher education.

Librado, P., Rozas, J. (2009) "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data", *Bioinformatics* 25, pp. 1451–1452.

Lindegaard, C. 1995. Classification of water-bodies and pollution, pp. 385–404. In P. D. Armitage, P. S. Cranston and L. C. V. Pinders (eds.), The Chironomidae: biology and ecology of non-biting midges. Springer Netherlands, Dordrecht.

Lourenco, A., Sequeira, E., Santovaia, H. and Gomes, C. (2014) "Magnetic geochemical and pedagogical characterization of soil profiles from different environments and geological backgrounds near Coimbra, Portugal", *Geoderma,* 213, pp. 408-418.

Loska, K., Wiechula, D. (2003) "Application of principal component analysis for the estimation of source of heavy metal contamination in surface sediments from the Rybnik Reservoir", *Chemosphere,* 51, pp. 723-733.

Lu, G, Zhu, L. and Yu, J. (2012) " Mineral magnetic properties of Chinese paddy soils and its pedogenic implications", Catena, 93, pp. 9-17.

Mandaville, S.M. (2002). *Benthic Macroinvertebrates in Freshwaters- Taxa Tolerance Values,*

Metrics, and Protocols. [online] Available at: http://www.chebucto.ns.ca/science/SWCS/H-1/tolerance.pdf

Marchant, R., Wells, F., Newall, P. (2000) " Assestment of an ecoregion approach for classifying macroinvertebrate assemblage from streams in Victoria, Australia", *Journal of the North American Benthological Society*, 19(3), pp. 497-500.

Mardulyn P, Milinkovitch MC (2005) Evaluating intra- specific ''network'' construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? Systematic biology, 54, 363–372

Marescaux, J., Doninck, K. (2013) "Using DNA barcoding to differentiate invasive *Dreissena* species (Mollusca, Bivalvia)", *ZooKeys*, 365, PP. 235-244.

Martin, S., Proulx, I., Hare, L. (2008) "Explaining metal concentrations in sympatric *Chironomus* species", *Limnology and Oceanography*, 53(2), pp. 411-419.

McKie, B. G., R. G. Pearson, and P. S. Cranston. 2005. Does biogeographical history matter? Diversity and distribution of lotie midges (Diptera: Chironomidae) in the Australian Wet Tropics. Austral Ecology 30:1-13.

Marsay, N., (2018) "Is the remediation at Parys Mpuntain Successfully Reducing Acid Mine drainage?". *Scientific Research Publishing*, 9.pp. 540-553.

Murdock, K., Wilkie, K., Brown, L. (2013) "Rock magnetic properties, magnetic susceptibility, and organic geochemistry comparison in core LZ1029-7 Lake El'gygytgyn, Russia Far East", *Climate of the Past*, 9, pp. 467-479.

Narum, S.R., Buerkle, C.A., Davey, J.W., Miller, M.R. and Hohenlohe, P.A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, [online] 22(11), pp.2841–2847.

Nyman, M., Korhola, A., Brooks, S. (2005) "The distribution and diversity of Chironomidae (Insecta: Diptera) in western Finnish Lapland, with special emphasis on shallow lakes", *Global Ecology and Biogeography*.

Orendt, C. (1999) “Chironomids as Bioindicators in acidified streams: a contribution to the acidify tolerance of Chironomid species with a classification in sensitivity classes”, *International. Rev. Hydrobiology*, (5), pp. 439-449.

Panatta, A., Stenert, C., Martins dos Santos, E., Maltchik, L. (2007) “Diversity and Distribution of Chironomid Larvae in Wetlands in Southern Brazil”, *Journal of the Kansas Entomological Society*, 80(3), pp. 229-242.

Panda UC, Sundaray SK, Rath P, Nayak BB, Bhatta D (2006) Application of factor and cluster analysis for characterization of river and estuarine water systems—a case study:MahanadiRiver (India). J Hydrol 331:434–445.

Peplow D (1999). Environmental Impacts of Mining in Eastern Was- hington, Center for Water and Watershed Studies Fact Sheet, Univer- sity of Washington, Seattle.

Pinder, L. (1986) “Biology of freshwater Chironomidae.”, *Annual Review of Entomology*, 31, pp.1–23.

Pongor, S. and Landsman, D. (1999). Bioinformatics and the developing world. *Biotechnology and development monitor*, [online] 40, pp.10–13. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4422384/ [Accessed 26 Aug. 2019].

Poonam, T., Tanushree, B., Sukalyan, C. (2013) "Water qaulity indices-important tools for water quality assesment: a review. International journal of advances in chemistry, 1 (1).

Punti, T., Rieradevall, M. and Prat, N. (2009) “Environmental factors, spatial variation, and specific requirements of Chironomidae in Mediterranean reference streams”, *Journal of the North American Benthological Society*, 28(1), pp.247-265.

Rainbow, P. S. (2002) “Trace metal concentrations in aquatic invertebrates: Why and so what?”, Environ. *Pollut*., 120, pp.497–507.

Raunio, J., Heino, J. and Paasivirta, L. (2011). Non-biting midges in biodiversity conservation and environmental assessment: Findings from boreal freshwater ecosystems. *Ecological Indicators*,

[online] 11(5), pp.1057–1064.

Rios, C., Williams, C., Roberts, C. (2007). Removal for heavy metals from acid mine drainage (AMD) using coal fly ash, natural clinker and synthetic xeolites. *Journals of Hazardous Material.*

Rosenberg, D. M. and V. H. Resh. 1993. Freshwater biomonitoring and ben- thic macroinvertebrates. Chapman and Hall, New York, NY.

S.L. Pimm, G.L. Russell, J.L. Gittleman, T.M. Brooks, The future of biodiversity, Science 269 1995 347–350.

Sari, A., Duran, M., Sen, A. and Bardakci, F. (2015) "Investigation of Chironomidae (Diptera) relationships using mitochondrial COI gene",*Biochemical Systematics and Ecology,*59, pp.229-238.

Shadrin, N., Anufriieva, E., Belyakov, V., Bazhora, A. (2017) " Chironomidae larvae in hypersaline waters of the Crimea: diversity, distribution, abudance and production", *The European Zoological Journal*, pp. 61-72.

Shokralla, S., SpalL, J.L., GIbson, J.F. and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, [online] 21(8), pp.1794–1805.

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environmental Modelling and Software, 22, 464–475.

Skubala P., Kafel, A. (2004) "Oribatid mite communities and metal bioaccumulation in oribatid species (Acari, Oribatida) along the heavy metal gradient in forest ecosystems" *Environ.Pollut*. 132, pp.51–60.

Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). PNAS 103, 3657–366.

Sogin ML, Morrison HG, Huber JA et al. (2006) Microbial diversity in the deep sea and the underexplored 'rare bio- sphere'. Proceedings of the National Academy of Sciences, USA, 103.

Sweeney, B. W., Battle, J., Jackson, J., Dapkey, T. (2011) “Can DNA barcodes of stream macroinvertebrates improve descriptions of com- munity structureand water quality?”, J. *North Am. Benthol. Soc*., 30, pp.195–216.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends in Ecology & Evolution, 18, 70– 74.

Thompson, J. Higgins, D., Gibson, T. (1994) “CLUSTAL W: improving the sensitivity of progressive sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice”, *Nucleic Acids Res*., 22, pp.4673-4680.

Tonkin, T., Midgley, N. and Graham, D. (2014). *The potential of small unmanned aircraft systems and structure-from-motion for topographic surveys: A test of emerging integrated approaches at Cwm Idwal, North Wales.*

Townsend, C., Hildrew, G., Francis, J. (1983) “Community structure in some southern English streams: the influence of physicochemical factors”, *Freshw. BioI*., 13, PP. 521-544.

Traister, E.M., McDowell, W.H., Krám, P., Fottová, D. and Kolaříková, K. (2013). Persistent effects of acidification on stream ecosystem structure and function. *Freshwater Science*, [online] 32(2), pp.586–596.

Valentini A, Pompanon F, Taberlet P (2009b) DNA barcoding for ecologists. Trends in Ecology and Evolution, 24, 110–117.

Wilson, B., Pyatt, F. (2006) “Heavy metal dispersion, persistence, and bioaccumulation around an ancient copper mine situated in Anglesey, UK”, *Ecotoxicology and environmental Safety*, 66, pp. 223-231.

Wilkerson, R.C., Linton, Y.M., Fonseca, D.M., Schultz, T.R., Price, D.C. and Strickman, D.A., 2015. Making mosquito taxonomy useful: a stable classification of tribe Aedini that balances utility with current knowledge of evolutionary relationships. PLoS One, 10(7).

Wright, S., (1951) “The genetical structure of populations”, *Ann.Eugen*, 15, pp. 323–354.

Yang, P., Byrne, J., Yang, M. (2016) " Spatial variability of soil magnetic susceptibility, organic carbon and total nitrogen from farmland in northern China", *Catena,* 145, pp. 92-98.

Yuan, J., Zhu, Q. and Liu, B., 2014. Phylogenetic and Biological Significance of Evolutionary Elements from Metazoan Mitochondrial Genomes. *PLoS ONE,* 9(1), p.e84330.

Zinchenko, T., Golovatyuk, V. (2012) "Salinity Tolerance of Macroinvertebrates in Stream Waters (Review)", *Aridnye Ekosistemy,* 19(3), pp. 5-17.

Glossary

Chironomidae: Insect family of true flies that includes the non-biting midges.

Diptera: An order of two-winged insects also known as true flies.

DNA-barcoding: Method of species identification using a short section of DNA.

MtDNA: Mitochondrial DNA.

CO1: Cytochrome oxidase 1 also known as Cox1.Mitochondrial DNA.

Genus: A taxonomic category that ranks above species and below family.

µL: Microlitre. SI measurement of liquids equal to 10^{-6} litres..

BioEdit: Software used for sequence alignment editing and analysis.

BLAST: Basic Local Alignment Search Tool. Bioinformatics program for searching DNA sequences.

Bray-Curtis similarity: A statistic used to quantify the similarity between two different sites, based on counts at each site.

Clustal Wallace: Bioinformatic computer programme used for multiple sequence alignments

% Identity: Percentage Identity. The similarity between two sequences.

E value: A parameter used to represent the number of matches (hits) that can be expected to be seen by chance.

GenBank: Public online database consisting of nucleotide sequences. Part of the NCBI.

PRIMER: A statistical package that is a collection of specialist univariate, multivariate, and graphical routines for analyzing species sampling data for community ecology.

Haplotype diversity (Hd): The probability that two randomly sampled alleles are different.

Max Score: The highest alignment score of a set of aligned segments from the same subject (data base) sequence.

Query Cover: The percentage of a given input sequence (the query) that has aligned with a

sequence

Total Score: The sum of alignment scores of all segments from the same subject sequence.

NCBI: The National Center for Biotechnology Information.

Phylogenetic network: Any graph used to visualise evolutionary relationships based on networks.

Neighbour-Joining: A clustering method for the creation of phylogenetic trees.

Nucleotide sequences: A sequence of letters (ACGT) that express the order of nucleotides that form alleles within a DNA molecule.

Nucleotide diversity (π): The average number of nucleotide differences per site in pairwise comparisons among DNA sequences.

Polymerase Chain Reaction (PCR): A molecular biology method used to make copies of a specific DNA region.

Simpson's diversity index: Calculation used in ecological surveys to calculate the diversity of species and their abundance present in a given area.

Heavy Metals: Any metallic chemical element that has a relatively high density and is toxic or poisonous at low concentrations.

ICP-OES: Inductively coupled plasma - optical emission spectrometry, is a technique in which the composition of elements in (mostly water-dissolved) samples can be determined using plasma and a spectrometer.

MS: A dimensionless proportionality constant that indicates the degree of magnetization of a material in response to an applied magnetic field.

IRM: A *remanent magnetization* (RM) acquired without the aid of changes in temperature.

PCA: A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components.

FA: A statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors.

DA: A technique that is used by the researcher to analyze the research data when the criterion or the dependent variable is categorical and the predictor or the independent variable is interval in nature.

ANOSIM: A non-parametric statistical test widely used in the field of ecology. The test was first suggested by K. R. Clarke[1] as an ANOVA-like test, where instead of operating on raw data, operates on a ranked dissimilarity matrix.

MSD: A means of visualizing the level of similarity of individual cases of a dataset.

Appendix

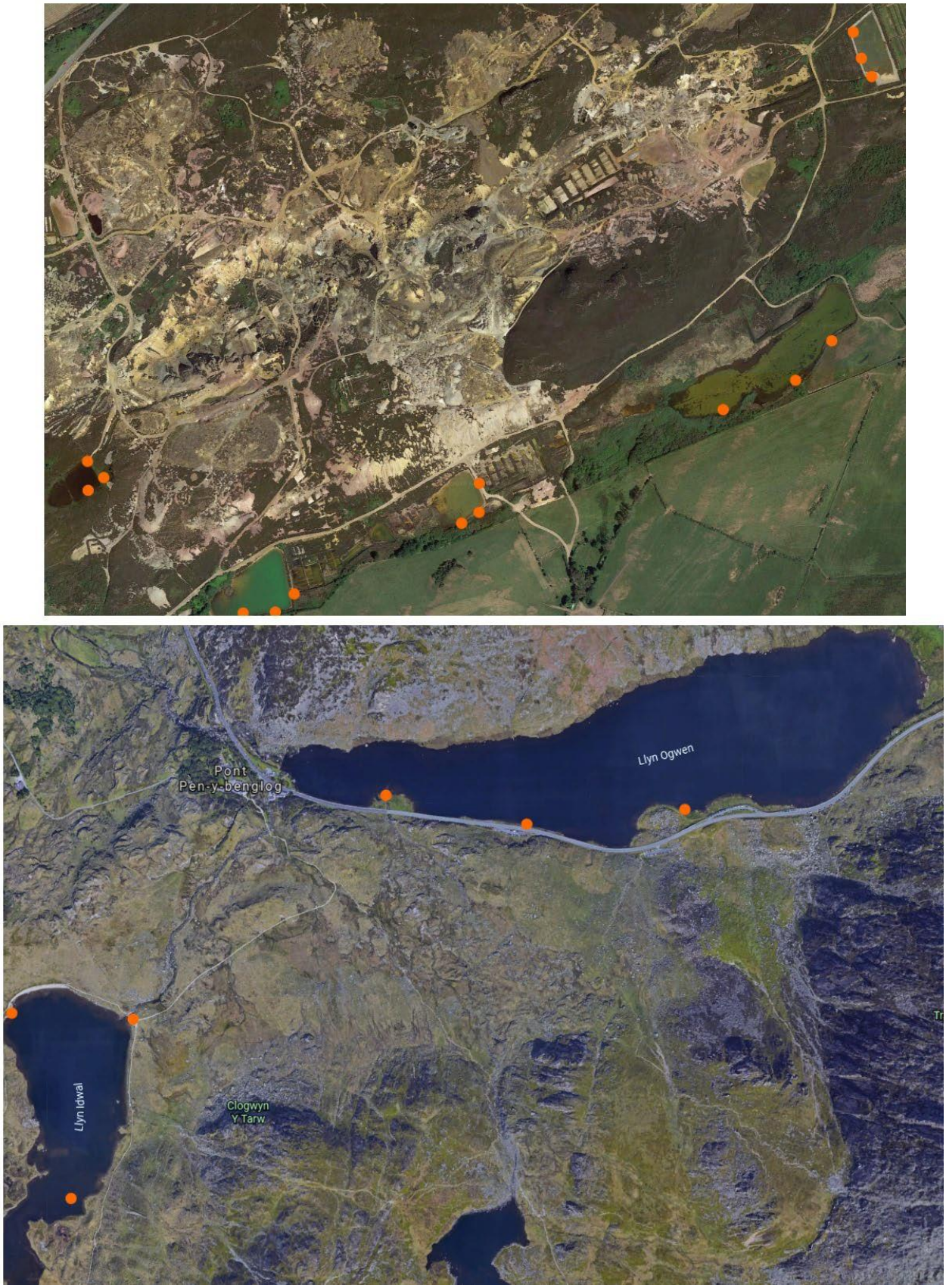


Figure S1 . Three sampling sites within the seven ponds in Wales.

Site ID	Latitude	Longitude
WAPM1	53.383925	-4.35002
WAPM2	53.382656	-4.346298
WAPM3	53.384043	-4.341278
WAPM4	53.385846	-4.334199
WAPM5	53.38941	-4.332357
WACI1	53.117967	-4.023639
WACI2	53.123072	-4.014895

Table S1. GPS coordinates of sampling sites.

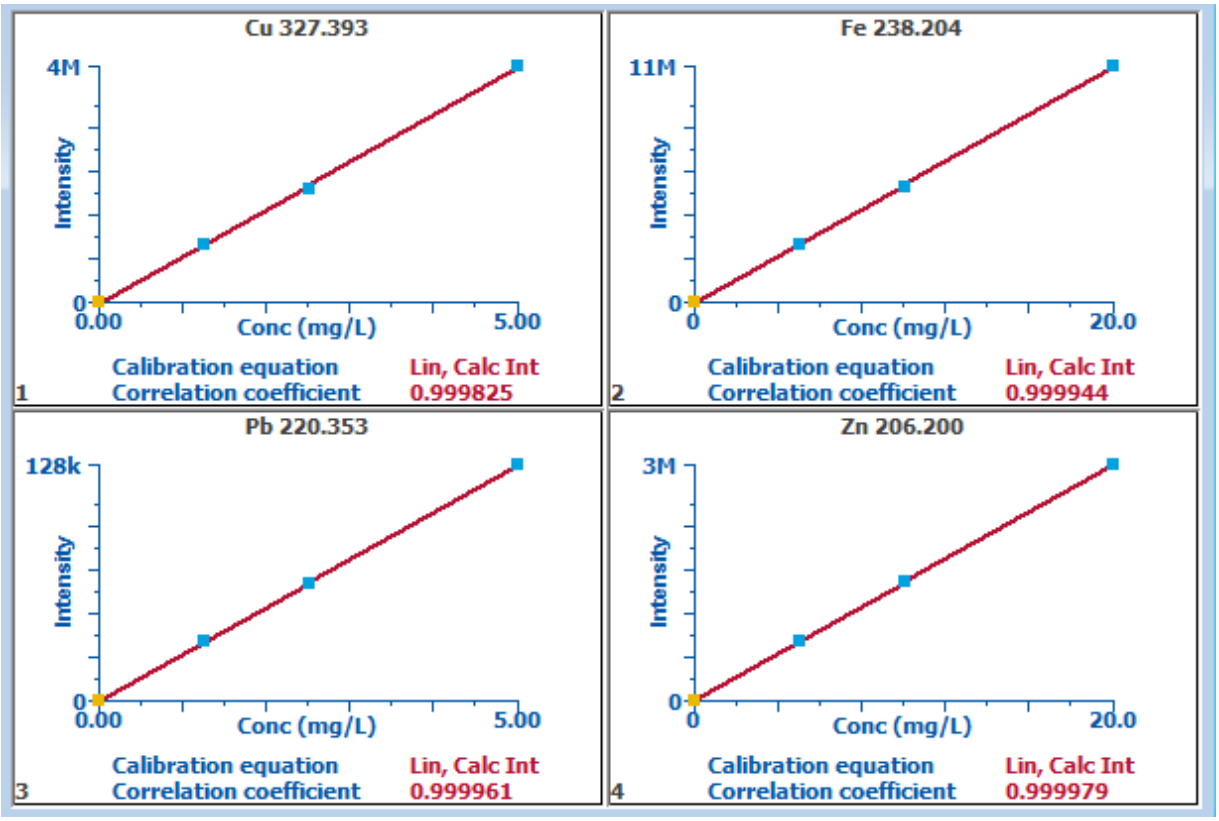


Figure S2 Calibration curve for Cu,Fe,Pb and Zn.

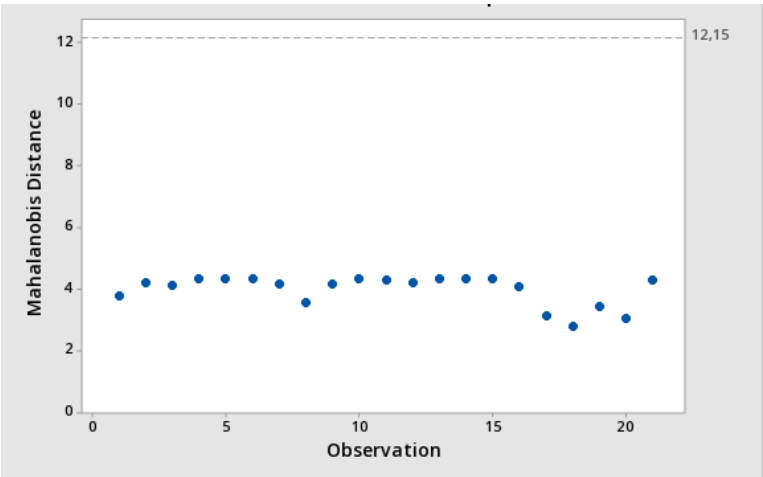


Figure S3. Outlier plot for Principal Component Analysis.

Table S2/ Eigenanalysis of the Correlation Matrix

Eigenanalysis of the Correlation Matrix																
Eigenvalue	70.9	32.98	21.77	13.11	0.9	0.7	0.5	0.4	0.2	0.2	0.1	0.1	0.0	0.0	0.0	0
Proportion	0.4	0.19	0.13	0.08	0.1	0	0	0	0	0	0	0.01	0	0	0	0
Cumulative	**0.4**	**0.61**	**0.74**	**0.82**	0.9	0.9	0.9	1	1	1	1	1	1	1	1	1

Table S3 Overall correct classification in Discriminant analysis

Overall No. Correct	Proportion
21	19	0.905

Table S4. Summary of Classification of the seven ponds in Wales where C1 and 2 rwpresent respectively pond 1 and 2 from Cwm-Idwal and pond 1-5 represent Ponds from Parys Mountain.

Summary of Classification							
Put in group	True Group						
	C1	C2	P1	P2	P3	P4	P5
C1	2	1	0	0	0	0	0
C2	1	2	0	0	0	0	0
P1	0	0	3	0	0	0	0
P2	0	0	0	3	0	0	0
P3	0	0	0	0	3	0	0
P4	0	0	0	0	0	3	0
P5	0	0	0	0	0	0	2
TOT No.	3	3	3	3	3	3	3
No. correct	2	2	3	3	3	3	3
Proportion	0.667	0.667	1	1	1	1	1

Table S5. Squared distance between groups in discriminant analysis.

Squared Distance Between Groups							
	C1	**C2**	**P1**	**P2**	**P3**	**P4**	**P5**
C1	0	4	3568	22435	102174	162524	109113
C2		0	3640	22139	101562	161783	108439
P1			0	23638	113495	176770	129617
P2				0	36628	75802	50942
P3					0	7202	6153
P4						0	11890
P5							0

Table S6. Summary of Misclassified Observation.

Summay of Misclassified Observations					
Observation	**True Group**	**Predicted Group**	**Group**	**Squared Distance**	**Probability**
16**	C1	C2	C1	8	0.12
			C2	4	0.88
			P1	3468	0
			P2	22024	0
			P3	101723	0
			P4	162024	0
			P5	108833	0
19**	C2	C1	C1	2	0.713
			C2	4	0.284
			P1	3659	0
			P2	22400	0
			P3	102022	0
			P4	162336	0
			P5	108855	0

Figure S4A. Zoomed in section of the phylogenetic tree.

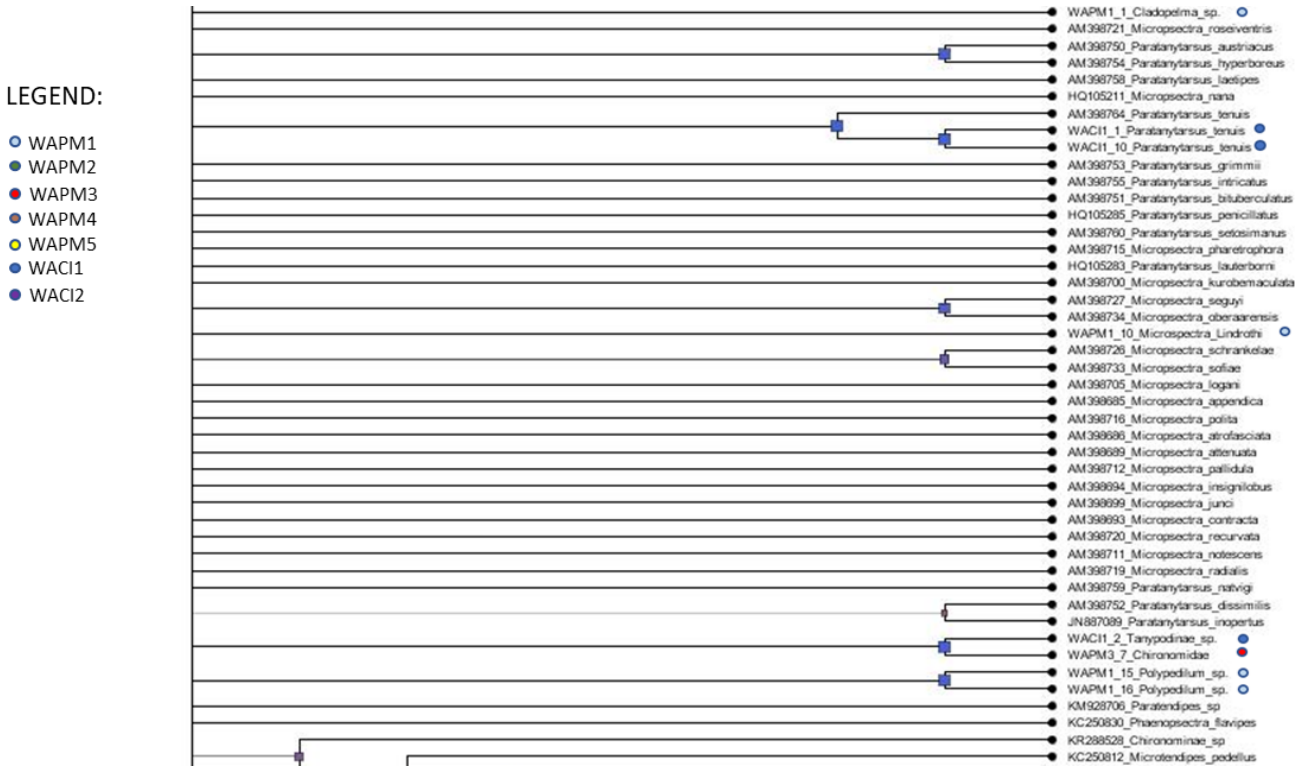


Figure S4B. Zoomed in section of the phylogenetic tree.

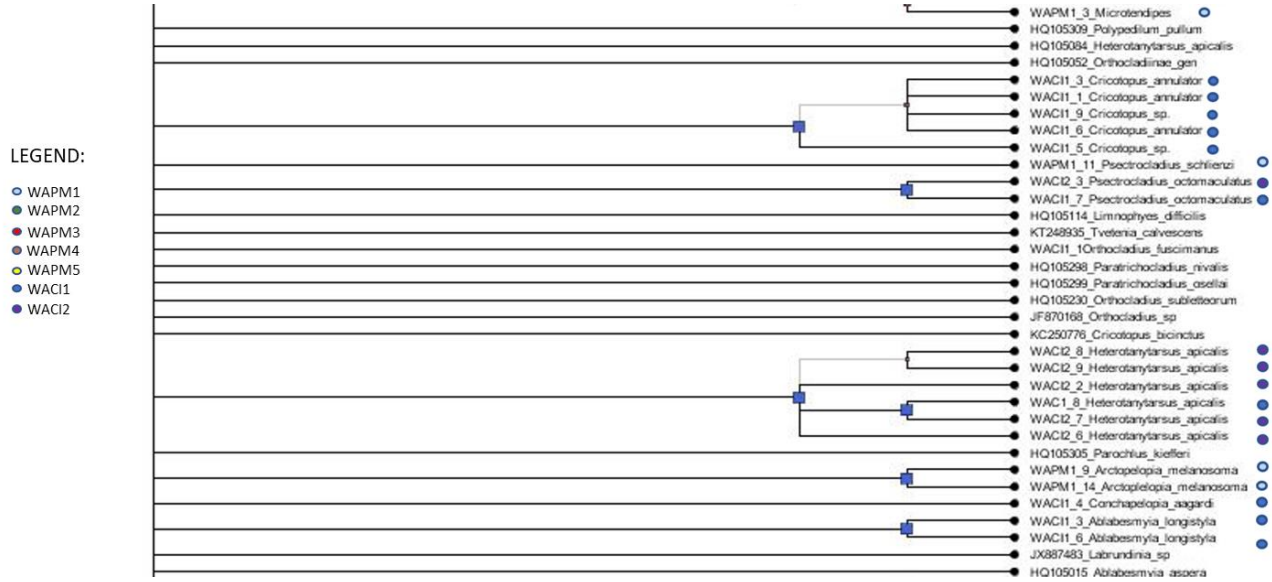


Figure S4C. Zoomed in section of the phylogenetic tree.

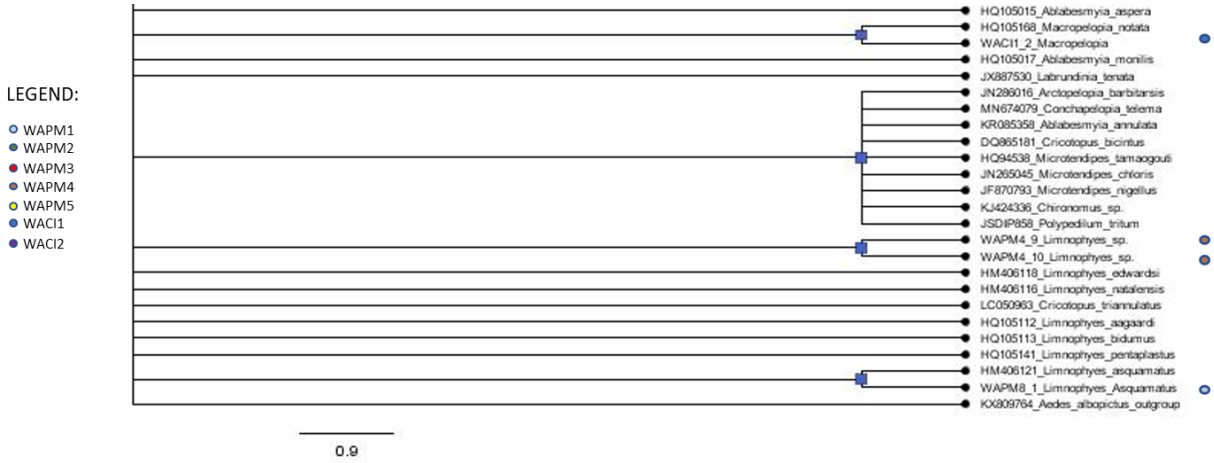


Figure S4D
. Zoomed in section of the phylogenetic tree.